

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:44:37 ; Search time 41 Seconds

(without alignments)
677.491 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175

Perfect score: 923

Sequence: 1 HTHQDFQPVHLVLAINTPLS.....RLLEQKAASCHNSIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	923	100.0	184	22	Murine endostatin
2	923	100.0	191	21	Murine endostatin
3	923	100.0	191	23	Amino acid sequenc
4	923	100.0	207	23	Mouse endostatin
5	920	99.7	183	23	Murine endostatin p
6	920	99.7	184	20	Murine endostatin
7	920	99.7	184	21	Murine angiogenesis
8	920	99.7	184	23	Human endostatin p
9	920	99.7	207	22	Murine endostatin

10	920	99.7	207	22	AA071930	Murine endostatin
11	920	99.7	218	20	AA086691	Murine gene therap
12	920	99.7	580	20	AA086692	Murine gene therap
13	920	99.7	684	20	AA25114	Mouse alpha-1 coll
14	920	99.7	1288	18	AAW26328	Mouse alpha-1 coll
15	920	99.7	1288	20	AAW92297	Endostatin protein
16	918	99.5	184	20	AAV18409	Endostatin protein
17	874	94.7	185	20	AAU06197	Anti-angiogenic en
18	813	88.1	180	22	AAU08899	Human Endostatin(T
19	813	88.1	181	22	AAU08898	Human Endostatin(T
20	813	88.1	182	21	AAU28339	Human endostatin
21	813	88.1	182	21	AAU94323	Human endostatin p
22	813	88.1	182	21	AAU59622	Human Endostatin(T
23	813	88.1	182	22	AAU00897	Amino acid sequenc
24	813	88.1	182	23	AAU77951	Human endostatin p
25	813	88.1	183	20	AAU08693	Human endostatin p
26	813	88.1	183	20	AAU02113	Amino acid sequenc
27	813	88.1	183	21	AAU30493	Human endostatin p
28	813	88.1	183	21	AAU16451	Human endostatin p
29	813	88.1	183	21	AAU90771	Human angiogenesis
30	813	88.1	183	21	AAU70252	Human angiogenesis
31	813	88.1	183	22	AAU00896	Human Endostatin(T
32	813	88.1	183	22	AAU49379	Human endostatin S
33	813	88.1	183	23	AAU79901	Human endostatin p
34	813	88.1	183	23	AAU97132	Human endostatin
35	813	88.1	183	23	AAU49503	Human endostatin p
36	813	88.1	183	23	AAU48895	Human endostatin
37	813	88.1	183	24	AAU79753	Murine endostatin
38	813	88.1	184	23	ABG31794	Human HMW endostat
39	813	88.1	193	21	AAU90877	Human HMW endostat
40	813	88.1	195	21	AAU90874	Amino acid sequenc
41	813	88.1	216	21	AAU30495	Synthetic plasmid
42	813	88.1	275	23	AAU76689	Human collagen XVI
43	813	88.1	310	23	AAU76688	Human Endostatin/I
44	813	88.1	513	23	ABG73586	Human ovarian anti
45	813	88.1	682	23	ABP41878	

ALIGNMENTS

RESULT 1

AA049380
ID AAB49380 standard; Protein; 184 AA.

AC AAB49380;

DT 02-MAR-2001 (first entry)

DE Murine endostatin SEQ ID NO: 4.

KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;
cancer; inflammation; angiogenesis-dependent disease.

OS Mus musculus.

PN WO200067771-A1.

PD 16-NOV-2000.

PF 02-MAY-2000; 2000WO-US12063.

PR 06-MAY-1999; 59US-0132907.

PR 14-JUL-1999; 59US-0353333.

XX (BURN-) BURNHAM INST.

XX Vuori K;

XX WPI; 2001-040937/05.

XX N-PSDB; AAC88290.

XX Endostatin peptide comprising at least four endostatin amino acid

PT residues are e.g. angiogenesis inhibitors for treating cancer and
 PT diabetic retinopathy -
 XX
 PS Disclosure; Fig 1; 146pp; English.
 XX
 CC The present invention provides endostatin peptides which can be used in
 CC the modulation of angiogenesis. This is useful in the treatment of
 CC cancers, inflammation, rheumatoid arthritis, chronic articular
 CC rheumatism, psoriasis, disorders associated with inopportune invasion of
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
 CC of prematurity, macular degeneration, corneal graft rejection,
 CC retrolental fibroplasia, rubeosis, capillary proliferation in
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
 CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque
 CC neovascularisation, telangiectasia, haemophiliac joints and wound
 CC granulation. In addition, the peptides can be used as birth control
 CC agents.
 XX
 SQ Sequence 184 AA;
 Query Match 100.0%; Score 923; DB 22; Length 184;
 Best Local Similarity 100.0%; Pred. No. 2.8e-102;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HTHQDFQPVHLVALNTPLSGMGRGADFCQFQARAVGLSGTFRAFLLSRLODLYSI 60
 DB 1 HTHQDFQPVHLVALNTPLSGMGRGADFCQFQARAVGLSGTFRAFLLSRLODLYSI 60
 QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLHRHPAPQKSVW 120
 DB 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLHRHPAPQKSVW 120
 QY 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIE 175
 DB 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIE 175
 RESULT 2
 AAB28398
 ID AAB28398 standard; Protein; 191 AA.
 XX
 AC AAB28398;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Murine endostatin.
 XX
 KW Murine; endostatin; cytostatic; antiproliferative;
 KW vascular endothelial growth factor; VEGF; antibody; VEGF2 receptor;
 KW cancer; vascularised solid tumour.
 XX
 OS Mus sp.
 XX
 PN WO200064946-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 28-APR-2000; 2000WO-US11367.
 XX
 PR 28-APR-1999; 99US-0131432.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Thorpe PE, Brekken RA;
 XX
 DR WPI; 2000-687317/67.
 DR N-PSDB; AAC67777.
 XX
 PT Immunogenic composition for the treatment and diagnosis of cancer
 PT comprises an anti-VEGF (vascular endothelial growth factor) antibody
 PT binding the same epitope as the monoclonal antibody ATCC PTA 1595 -
 XX
 PS Example 10; Page 290-291; 298pp; English.

XX
 CC The present invention relates to anti-Vascular Endothelial Growth Factor
 CC (VEGF) antibodies that bind to the same epitope as the monoclonal
 CC antibody ATCC PTA 1595 and which significantly inhibit VEGF binding to
 CC the VEGF receptor VEGFR2, without inhibiting VEGF binding to the VEGF
 CC receptor VEGFR1. The present sequence is murine endostatin. Endostatin
 CC may be conjugated onto the anti-VEGF antibodies of the present invention.
 CC The anti-VEGF antibodies of the present invention are useful for the
 CC treatment and diagnosis of cancer, especially vascularised solid tumours.
 XX
 SQ Sequence 191 AA;
 Query Match 100.0%; Score 923; DB 21; Length 191;
 Best Local Similarity 100.0%; Pred. No. 2.9e-102;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HTHQDFQPVHLVALNTPLSGMGRGADFCQFQARAVGLSGTFRAFLLSRLODLYSI 60
 DB 8 HTHQDFQPVHLVALNTPLSGMGRGADFCQFQARAVGLSGTFRAFLLSRLODLYSI 67
 QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLHRHPAPQKSVW 120
 DB 68 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGRDVLHRHPAPQKSVW 127
 QY 121 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIE 175
 DB 128 HGSDPSGRRLMESYCETWRTTGTGATGQASSLLSGRLLEOKAASCHNSYIVLCIE 182
 RESULT 3
 AAU77950
 ID AAU77950 standard; Protein; 191 AA.
 XX
 AC AAU77950;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Amino acid sequence for mouse endostatin.
 XX
 KW Mouse; immunoconjugate; anti-vascular endothelial growth factor antibody;
 KW anti-VEGF antibody; monoclonal antibody 2C3 ATCC PTA 1595; VEGF receptor;
 KW VEGFR2; KDR/Flk-1; VEGFR1; Flt-1; angiogenesis; macular degeneration;
 KW ocular neovascular disease; cancer; vascularised solid tumour; AIDS;
 KW metastatic tumour; endothelial cell proliferation; inflammatory disorder;
 KW atherosclerosis; diabetic retinopathy; corneal graft rejection;
 KW acquired immune deficiency syndrome; infection; restenosis; fungal ulcer;
 KW sickle cell anaemia; endometriosis; endostatin.
 XX
 OS Mus sp.
 XX
 PN AU200179401-A.
 XX
 PD 06-DEC-2001.
 XX
 PF 12-OCT-2001; 2001AU-0079401.
 XX
 PR 28-APR-2000; 2000AU-0048049.
 PR 12-OCT-2001; 2001AU-0079401.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Thorpe PE, Brekken RA;
 XX
 DR WPI; 2002-281368/33.
 DR N-PSDB; ABK47719.
 XX
 PT Immunoconjugate compositions for treating cancer by inhibiting
 PT angiogenesis and for delivering a diagnostic agent to tumour, comprises
 PT anti-vascular endothelial growth factor antibody attached to a
 PT biological agent -
 XX
 PS Example 10; Page 11-12 (Sequence listing); 300pp; English.

CC The present invention relates to antibody-based compositions comprising
CC an immunocombination such as anti-vascular endothelial growth factor
CC (VEGF) antibody (Ab) (or its antigen-binding fragment), attached to a
CC biological agent, where the Ab binds to the same epitope as the
CC monoclonal antibody (Mab) 2C3 ATCC PTA 1595, and significantly inhibits
CC VEGF binding to the VEGF receptor VEGFR2 (KDR/Flk-1) without inhibiting
CC VEGF binding to the VEGF receptor VEGFR1 (Flt-1). The compositions
CC of the invention are useful in therapy, and diagnosis, for inhibiting
CC angiogenesis in an animal having ocular neovascular disease or macular
CC degeneration, and for delivering a biological agent to a vascularised
CC tumour. The compositions can also be used for treating cancer and
CC subjects at risk of developing, a vascularised solid tumour, a metastatic
CC tumour or metastases from a primary tumour. The composition is useful
CC for specifically inhibiting VEGF-induced endothelial cell proliferation,
CC without significantly inhibiting VEGF-induced macrophage, osteoclast or
CC chondrocyte function. The compositions can be used for treating various
CC diseases such as inflammatory disorders, atherosclerosis, diabetic
CC retinopathy, retinosis, acquired immune deficiency syndrome (AIDS),
CC blood borne tumours, corneal graft rejection, Crohn's disease, fungal
CC ulcers, infections, sickle cell anaemia, and endometriosis. The present
CC sequence represents mouse endostatin. Endostatin may be attached or
CC functionally associated with anti-VEGF antibodies.

XX SQ Sequence 191 AA;

Query Match 100.0%; Score 923; DB 23; Length 191;
Best Local Similarity 100.0%; Pred. No. 2.9e-102;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
DB 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDVLHPAMPQKSVW 120
DB 68 VRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDVLHPAMPQKSVW 127
QY 121 HGSDPSGRRLMESYCYETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
DB 128 HGSDPSGRRLMESYCYETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 182

RESULT 4
ABB79902
ID ABB79902 standard; Protein; 207 AA.

XX AC ABB79902;

XX DT 05-DEC-2002 (first entry)

XX DE Mouse endostatin.

XX KW Endostatin; mouse; ophthalmological; ocular neovascularisation;
XX KW choroidal neovascularisation; gene therapy.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

XX FT Misc-difference 117 /note= "encoded by GTG"

XX FT WO200267971-A2.

XX FT 06-SEP-2002.

XX FF 21-FEB-2002; 2002WO-US05336.

XX PR 22-FEB-2001; 2001US-270787P.

XX PR 04-APR-2001; 2001US-281296P.

XX PA (NOVS) NOVARTIS AG.

XX PI Brazzell RK, Campochiaro PA, Dixon KH;

XX DR WPI; 2002-698636/75.

XX DR N-PSDB; ABQ81194.

XX PT Treating or preventing choroidal neovascularization comprises
XX PT increasing the amount of endostatin in ocular tissues of afflicted
XX PT individuals to a choroidal neovascularization inhibiting level -
XX PS Disclosure; Page 40; 44pp; English.

XX CC The present sequence is the protein sequence of murine endostatin
XX CC plus the murine Ig kappa leader sequence. A claimed method for the
XX CC treatment of ocular neovascularisation, especially choroidal
XX CC neovascularisation, involves increasing the level of endostatin in
XX CC ocular tissue, especially the level of human endostatin (see
XX CC ABB79901), its fragment, derivative or variant. The increase is
XX CC effected by administering a viral vector, especially an adenovirus,
XX CC adeno-associated virus, a retrovirus or lentivirus vector,
XX CC comprising an endostatin-encoding nucleic acid. Cells secreting
XX CC endostatin may be encapsulated and implanted within an individual.
XX CC The method is used when ocular neovascularisation is caused by
XX CC histoplasmosis, pathological myopia, angiod streaks, anterior
XX CC ischaemic optic neuropathy, bacterial endocarditis, Best's disease,
XX CC birdshot retinopathy, choroidal haemangioma, choroidal
XX CC naevi, choroidal nonperfusion, choroidal osteomas, choroidal
XX CC rupture, choroideraemia, chronic retinal detachment, coloboma of
XX CC the retina, drusen, endogenous Candida endophthalmitis, extrapapillary
XX CC hamartoma of the retina, fundus flavimaculatus, idiopathic, macular hole, malignant
XX CC melanoma, membranoproliferative glomerulonephritis (type II),
XX CC metallic intraocular foreign body, morning glory disc syndrome,
XX CC multiple evanescent white-dot syndrome, neovascularisation of ora
XX CC serrata, operating microscope burn, optic nerve head pits,
XX CC photocoagulation, punctate inner chorioidopathy, rubella,
XX CC sarcoidosis, seriginous or geographic chorioiditis, subretinal
XX CC fluid drainage, tiled disc syndrome, Toxoplasma retinochoroiditis,
XX CC tuberculosis, Vogt-Koyanagi-Harada syndrome, diabetic retinopathy,
XX CC non-diabetic retinopathy, brain vein occlusion, central retinal
XX CC vein occlusion, retinopathy in premature infants, rubeosis iridis,
XX CC neovascular glaucoma, perifoveal telangiectasis, sickle cell
XX CC retinopathy, Sall's disease, retinal vasculitis, von Hippel
XX CC Lindau disease, radiation retinopathy, retinal cryoinjury,
XX CC retinitis pigmentosa, retinochoroidal coloboma, corneal
XX CC neovascularisation due to herpes simplex keratitis, corneal ulcers,
XX CC keratoplasty, pterygia and trauma (all claimed).

XX SQ Sequence 207 AA;

Query Match 100.0%; Score 923; DB 23; Length 207;
Best Local Similarity 100.0%; Pred. No. 3.3e-102;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
DB 24 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 83
QY 61 VRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDVLHPAMPQKSVW 120
DB 84 VRADRGSPVIVNLKDEVLSPSWDSLFSGSQGLQPGARIFSPDGRDVLHPAMPQKSVW 143
QY 121 HGSDPSGRRLMESYCYETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
DB 144 HGSDPSGRRLMESYCYETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 198

RESULT 5

XX AAM49504

XX ID AAM49504 standard; Protein; 183 AA.

XX AC AAM49504;

XX XX 07-MAY-2002 (first entry)

DE Mouse endostatin protein.
XX Endostatin; murine; proliferation; blood vessel endothelium;
KW regeneration; tumour; blood vessel; treatment; amplification.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FT Misc-difference 181
FT /label= Ser, Phe
XX
XX CN1177005-A.
XX
XX 25-MAR-1998.
XX
XX 10-SEP-1997; 97CN-0107112.
XX
XX 10-SEP-1997; 97CN-0107112.
XX
XX (XUGG/) XU G.
XX Xu G, Ren M, Xu L;
XX WPI; 2002-106746/15.
XX
XX Gene clone of inhibitory factor for hyperplasia of inner blood vessel
PT cells in human body's real tumor, and its use in anti-tumor blood
PT vessel regeneration -
XX
XX Disclosure; Page 4 (Disclosure); 6pp; Chinese.
XX
XX This invention describes a novel preparation which inhibits the
CC proliferation of blood vessel endothelium and prevents the regeneration
CC activity of tumour blood vessels. The preparation can also be used as a
CC biological preparation in the treatment of tumours. This sequence
CC represents the murine endostatin protein described in the invention.
XX
XX Sequence 183 AA;
SQ
Query Match 99.7%; Score 920; DB 23; Length 183;
Best Local Similarity 99.4%; Pred. No. 6.3e-102;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HTHQDFQVPLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HTHQDFQVPLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Qy 61 VRRADRGSPVIVNLKDEVLSFSGQGLQPGARIFSPDGRDLRHPAPQKSVW 120
Db 1 HTHQDFQVPLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Qy 61 VRRADRGSPVIVNLKDEVLSFSGQGLQPGARIFSPDGRDLRHPAPQKSVW 120
Db 61 VRRADRGSPVIVNLKDEVLSFSGQGLQPGARIFSPDGRDLRHPAPQKSVW 120
Qy 121 HGSDPSGRRLMESYCETWRTTTCATGQASSLLSRLLEKQKAAACHNSYIVLCIE 175
Db 121 HGSDPSGRRLMESYCETWRTTTCATGQASSLLSRLLEKQKAAACHNSYIVLCIE 175
RESULT 6
AAY08689
ID AAY08689 standard; Protein; 184 AA.
AC AAY08689;
XX
XX 10-AUG-1999 (first entry)
XX
XX Murine endostatin protein fragment.
XX
XX Plasminogen; murine; angiostatin; endostatin; gene therapy; vector;
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina.
XX
XX Mus sp.
XX
XX WO9926480-A1.
PN

XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-US24950.
XX
XX 20-NOV-1997; 97US-0975424.
XX
XX (GENE-) GENETIX PHARM INC.
XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Bachelot T, Leboulch P, Pawliuk RJ;
XX
XX WPI; 1999-357696/30.
XX N-PSDB; AAX77715.
XX
XX Anti-angiogenic gene therapy vectors
XX
XX Disclosure; Fig 6; 83pp; English.
XX
XX This invention describes a novel viral gene therapy vector comprising a
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
CC from human or murine angiostatin, human or murine endostatin and
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
CC sufficiently attenuated for use in human gene therapy. The products of
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
CC ophthalmological activity. The vector is used in gene therapy for
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
CC expresses an anti-angiogenic polypeptide. An additional use comprises
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
CC inhibits angiogenesis in the vicinity of the retina. The vector is
CC administered to cells ex vivo and then administered to the patient.
XX
XX Sequence 184 AA;
SQ
Query Match 99.7%; Score 920; DB 20; Length 184;
Best Local Similarity 99.4%; Pred. No. 6.4e-102;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HTHQDFQVPLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HTHQDFQVPLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Qy 61 VRRADRGSPVIVNLKDEVLSFSGQGLQPGARIFSPDGRDLRHPAPQKSVW 120
Db 61 VRRADRGSPVIVNLKDEVLSFSGQGLQPGARIFSPDGRDLRHPAPQKSVW 120
Qy 121 HGSDPSGRRLMESYCETWRTTTCATGQASSLLSRLLEKQKAAACHNSYIVLCIE 175
Db 121 HGSDPSGRRLMESYCETWRTTTCATGQASSLLSRLLEKQKAAACHNSYIVLCIE 175
RESULT 7
AAY70258
ID AAY70258 standard; Protein; 184 AA.
AC AAY70258;
XX
XX 06-JUN-2000 (first entry)
XX
XX Murine angiogenesis inhibitor, endostatin.
XX
XX Murine; immunoglobulin Fc fragment; endostatin; immunofusin;
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;
KW antipsoriatic; anti-diabetic; ophthalmological; immunosuppressant;
KW vasotropic; vulnerary; treatment; antiarteriosclerosis; tumour;
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;
KW wound granulation; keloid scar; gene therapy.
XX
XX Mus musculus.
XX
XX WO200011033-A2.
PN

PR 25-OCT-1999; 99US-0433042.
ER 02-JUN-2000; 2000US-0586625.
XX
XX
FA (NOVS) NOVARTIS AG.
PA (SCRI) SCRIPPS RES INST.
XX
XX Barbas CF, Kadan M, Beerli R;
XX
XX WPI; 2001-308618/32.
DR N-PSDB; AAD06108.
XX
XX New fusion protein containing nucleotide-binding and ligand-binding
PT domains, useful e.g. in gene therapy of cancer, provides
PT ligand-activated control of gene expression -
XX
XX Example 19; Page 209; 218pp; English.
XX
XX The invention relates to fusion protein comprising a nucleotide-binding
CC domain (NBD), a ligand-binding domain (LBD) of an intracellular receptor
CC (ICR) and a transcription regulating domain (TRD). NBD is a polydactyl
CC zinc finger protein (ZFP), or a modular part of it, that interacts
CC specifically with a contiguous sequence of at least 3 nucleotides. The
CC fusion protein functions as a ligand-activated transcriptional regulator.
CC The fusion protein and the nucleic acid encoding it, are used to regulate
CC gene expression, particularly in gene therapy for treating malignant
CC cell proliferative diseases (e.g. colon cancer, prostate cancer,
CC renal-cell carcinoma) and non-malignant cell proliferative
CC diseases (e.g. psoriasis, pemphigus vulgaris, Behcet's syndrome and
CC lipid histiocytosis). The fusion protein and its DNA are also useful for
CC treating diseases caused by viruses in humans/plants, genetic and/or
CC acquired diseases. The fusion protein can be designed to target any
CC selected gene (endogenous or exogenous), and can be made to have
CC different selectivity or specificity for endogenous or exogenous ligands.
CC The present sequence is murine endostatin fused to an N-terminal
CC secretion signal. The corresponding cDNA sequence was used in the
CC construction of Left and shuttle plasmids containing regulatable
CC transgene cassettes for evaluation of Cys2-His2 Zinc finger DNA binding
CC domain (DBD)-Oestrogen receptor (ER) LBD regulators.
XX
XX
SQ Sequence 207 AA;
Query Match 99.7%; Score 920; DB 22; Length 207;
Best Local Similarity 99.4%; Pred. No. 7.6e-102;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 24 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 83
Qy 61 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSGDGRDLRHPAMPQKSVW 120
Db 84 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSGDGRDLRHPAMPQKSVW 143
Qy 121 HGSDFSGRRLMESYCETWRTTGTGATGQASLLSGRLLEKAAACHNSYIVLCIE 175
Db 144 HGSDFSGRRLMESYCETWRTTGTGATGQASLLSGRLLEKAAACHNSYIVLCIE 198
RESULT 10
AAB71930
ID AAB71930 standard; Protein; 207 AA.
XX
XX AAB71930;
XX
DT 10-MAY-2001 (first entry)
XX
XX Murine endostatin attached to Ig-kappa leader sequence.
DE
XX Mouse; endostatin; antitumour; cytostatic; antiarthritic; antipsoriatic;
KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;
KW lymphoma; leukaemia; immunoglobulin; Ig; Ig-kappa.

XX Mus sp.
XX WO200112830-A1.
XX
XX 22-FEB-2001.
XX
XX 11-AUG-2000; 2000WO-EP07865.
XX
XX 13-AUG-1999; 99US-0373938.
XX
XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Hallenbeck PL, Chen CT;
XX
XX WPI; 2001-202871/20.
DR N-PSDB; AAF60336.
XX
XX Adenoviral vector for treating tumors and disorders associated with
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA
PT sequence encoding an angiogenic inhibitor, particularly endostatin -
XX
XX Example 1; Fig 1B; 59pp; English.
XX
XX The nucleotide sequence encoding this protein was used in the
CC construction of an adenoviral vector which includes a DNA sequence
CC encoding endostatin. The adenoviral vector is useful for expressing
CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is
CC useful for treating other diseases and disorders associated with
CC angiogenesis, such as neovascular diseases of the eye, including diabetic
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The
CC vector inhibits, prevents or delays the growth of tumours by
CC preventing the formation of blood vessels in tumours, such as lymphoma
CC and leukaemia.
XX
XX Sequence 207 AA;
Query Match 99.7%; Score 920; DB 22; Length 207;
Best Local Similarity 99.4%; Pred. No. 7.6e-102;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 24 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 83
Qy 61 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSGDGRDLRHPAMPQKSVW 120
Db 84 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSGDGRDLRHPAMPQKSVW 143
Qy 121 HGSDFSGRRLMESYCETWRTTGTGATGQASLLSGRLLEKAAACHNSYIVLCIE 175
Db 144 HGSDFSGRRLMESYCETWRTTGTGATGQASLLSGRLLEKAAACHNSYIVLCIE 198
RESULT 11
AAY08691
ID AAY08691 standard; Protein; 218 AA.
XX
XX AAY08691;
XX
XX 10-AUG-1999 (first entry)
XX
XX Murine gene therapy peptide construct SP-Flag-Endo.
XX
XX Plasmidogen; murine; angiostatin; endostatin; gene therapy; vector;
KW anti-angiogenic; atenuation; cytostatic; anti-diabetic; ophthalmology;
KW tumour growth; solid tumour; diabetic retinopathy; retina; construct.
XX
XX Mus sp.
OS Synthetic.
XX

PN WO9926480-A1.
 XX 03-JUN-1999.
 XX 20-NOV-1998; 98WO-US24950.
 XX 20-NOV-1997; 97US-0975424.
 XX (GENE-) GENETIX PHARM INC.
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX Bachelot T, Leboulch P, Pawliuk RJ;
 XX WPI; 1999-357696/30.
 XX N-PSDB; AAX77717.
 XX Anti-angiogenic gene therapy vectors
 XX Example 1; Page 69; 83pp; English.
 XX This invention describes a novel viral gene therapy vector comprising a
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
 CC from human or murine angiostatin, human or murine endostatin and
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
 CC sufficiently attenuated for use in human gene therapy. The products of
 CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
 CC ophthalmological activity. The vector is used in gene therapy for
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
 CC expresses an anti-angiogenic polypeptide. An additional use comprises
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
 CC inhibits angiogenesis in the vicinity of the retina. The vector is
 CC administered to cells ex vivo and then administered to the patient.
 XX SQ Sequence 218 AA;
 Query Match 99.7%; Score 920; DB 20; Length 218;
 Best Local Similarity 99.4%; Pred. No. 8.2e-102;
 Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HTHQDFQPVHLHVALNTPLSGMRGIRGADFCQCFQARAVGLSGTFFAFLLSRLQDLYSI 60
 Db 35 HTHQDFQPVHLHVALNTPLSGMRGIRGADFCQCFQARAVGLSGTFFAFLLSRLQDLYSI 94
 QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLQPCGARIFFSDGRDVLHRHPAPQKSVW 120
 Db 95 VRRADRGSPVIVNLKDEVLSFSGQQLQPCGARIFFSDGRDVLHRHPAPQKSVW 154
 QY 121 HGSDFSGRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
 Db 155 HGSDFSGRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 209
 RESULT 12
 AAY08692
 ID AAY08692 standard; Protein; 580 AA.
 XX AC AAY08692;
 XX DT 10-AUG-1999 (first entry)
 XX DE Murine gene therapy peptide construct SP-K1-K2-K3-K4-Flag-Endo.
 XX KW Plasminogen; murine; angiostatin; endostatin; gene therapy; vector;
 XX anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;
 XX tumour growth; solid tumour; diabetic retinopathy; retina; construct.
 XX OS Mus sp.
 XX OS Synthetic.
 XX PN WO9926480-A1.
 XX 03-JUN-1999.

PF 20-NOV-1998; 98WO-US24950.
 XX 20-NOV-1997; 97US-0975424.
 XX (GENE-) GENETIX PHARM INC.
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX Bachelot T, Leboulch P, Pawliuk RJ;
 XX WPI; 1999-357696/30.
 XX N-PSDB; AAX77718.
 XX Anti-angiogenic gene therapy vectors
 XX Example 1; Page 72-74; 83pp; English.
 XX This invention describes a novel viral gene therapy vector comprising a
 CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen
 CC from human or murine angiostatin, human or murine endostatin and
 CC angiogenesis-inhibiting fusions and fragments, where the viral vector is
 CC sufficiently attenuated for use in human gene therapy. The products of
 CC the invention have anti-angiogenic, cytostatic, anti-diabetic and
 CC ophthalmological activity. The vector is used in gene therapy for
 CC inhibiting tumour growth in humans harbouring a solid tumour. The vector
 CC expresses an anti-angiogenic polypeptide. An additional use comprises
 CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide
 CC inhibits angiogenesis in the vicinity of the retina. The vector is
 CC administered to cells ex vivo and then administered to the patient.
 XX SQ Sequence 580 AA;
 Query Match 99.7%; Score 920; DB 20; Length 580;
 Best Local Similarity 99.4%; Pred. No. 3.4e-101;
 Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HTHQDFQPVHLHVALNTPLSGMRGIRGADFCQCFQARAVGLSGTFFAFLLSRLQDLYSI 60
 Db 397 HTHQDFQPVHLHVALNTPLSGMRGIRGADFCQCFQARAVGLSGTFFAFLLSRLQDLYSI 456
 QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLQPCGARIFFSDGRDVLHRHPAPQKSVW 120
 Db 457 VRRADRGSPVIVNLKDEVLSFSGQQLQPCGARIFFSDGRDVLHRHPAPQKSVW 516
 QY 121 HGSDFSGRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
 Db 517 HGSDFSGRRLMESYCETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 571
 RESULT 13
 AAY25114
 ID AAY25114 standard; Protein; 684 AA.
 XX AC AAY25114;
 XX DT 25-AUG-1999 (first entry)
 XX DE Mouse alpha1 (XVIII) collagen protein.
 XX KW Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;
 XX anti-angiogenic; heparin binding domain; receptor binding domain; mimic;
 XX alpha-helix A domain; carbohydrate recognition domain; CRD domain;
 XX treatment; angiogenesis; tumour; murine.
 XX OS Mus sp.
 XX PN WO9931616-A1.
 XX 24-JUN-1999.
 XX PF 16-DEC-1998; 98WO-US26783.
 XX 16-DEC-1997; 97US-0069727.

```

PA (HARD ) HARVARD COLLEGE.
XX
XX Hohenester E, Olsen BR, Sasaki T, Timpl R;
XX
XX WPI; 1999-395243/33.
XX
XX Identifying mimetics of mammalian endostatin
XX
XX Disclosure; Fig 5A-C; 75pp; English.
XX
XX This invention describes a novel method for identifying mimetics of
XX mammalian endostatin. The method comprises identifying a compound
XX having atomic coordinates with non-trivial similarity to selected
XX coordinates of atoms of a mammalian endostatin involves (a) providing
XX a library of atomic coordinates of compounds in a library of candidate
XX compounds, (b) comparing the library of atomic coordinates to the
XX selected coordinates of a mammalian endostatin and (c) selecting from the
XX library at least one candidate compound on the basis of selection
XX criteria which include similarities between the atomic coordinates of the
XX selected candidate compound and the atomic coordinates of the mammalian
XX endostatin. The invention also describes the use of an anti-angiogenic
XX fragment of endostatin comprising a domain selected from a heparin
XX binding domain, a receptor binding domain, and exposed on alpha-helix A
XX domain, and a carbohydrate recognition domain (CRD) domain. The methods
XX can be used for designing and selecting endostatin mimics. The compounds
XX identified can be used for treating undesired angiogenesis, e.g. tumours.
XX This sequence represents mouse alpha1(XVIII) collagen which is used in
XX the description of the method.
XX
XX Sequence 684 AA;
XX
XX Query Match          99.7%; Score 920; DB 20; Length 684;
XX Best Local Similarity 99.4%; Pred No. 4,4e-101;
XX Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVGLSGTFRFLSSRLQDLYSI 60
Db 502 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVGLSGTFRFLSSRLQDLYSI 561
QY 61 VERADRGSPVIVNLKDEVLSPSWDSLFSGSGOQOPGARIFFDGRDVLRRHPAMPQKSVW 120
Db 562 VERADRGSPVIVNLKDEVLSPSWDSLFSGSGOQOPGARIFFDGRDVLRRHPAMPQKSVW 621
QY 121 HGSDFSGRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 622 HGSDFSGRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 676

RESULT 14
AAW26328
ID AAW26328 standard; Protein; 1288 AA.
XX
XX AAW26328;
XX
XX DT 19-NOV-1997 (first entry)
XX
XX DE Mouse alpha-1 collagen (XVIII).
XX
XX KW Alpha-1 collagen; type XVIII collagen; cartilage degeneration.
XX
XX OS Mus musculus.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 303..308 /label= GXYGX'Y'_motif
XX FT Peptide 309..314 /label= GXYGX'Y'_motif
XX FT Peptide 315..320 /label= GXYGX'Y'_motif
XX FT Peptide 321..326 /label= GXYGX'Y'_motif
XX FT Peptide 337..342 /label= GXYGX'Y'_motif
XX FT Peptide /label= GXYGX'Y'_motif

```

```

FT Peptide 343..348 /label= GXYGX'Y'_motif
FT Peptide 349..354 /label= GXYGX'Y'_motif
FT Peptide 355..360 /label= GXYGX'Y'_motif
FT Peptide 361..366 /label= GXYGX'Y'_motif
FT Peptide 367..372 /label= GXYGX'Y'_motif
FT Peptide 373..378 /label= GXYGX'Y'_motif
FT Peptide 379..384 /label= GXYGX'Y'_motif
FT Peptide 385..390 /label= GXYGX'Y'_motif
FT Peptide 396..401 /label= GXYGX'Y'_motif
FT Peptide 402..407 /label= GXYGX'Y'_motif
FT Peptide 435..440 /label= GXYGX'Y'_motif
FT Peptide 441..446 /label= GXYGX'Y'_motif
FT Peptide 447..452 /label= GXYGX'Y'_motif
FT Peptide 453..458 /label= GXYGX'Y'_motif
FT Peptide 459..464 /label= GXYGX'Y'_motif
FT Peptide 470..475 /label= GXYGX'Y'_motif
FT Peptide 476..481 /label= GXYGX'Y'_motif
FT Peptide 482..487 /label= GXYGX'Y'_motif
FT Peptide 488..493 /label= GXYGX'Y'_motif
FT Peptide 494..499 /label= GXYGX'Y'_motif
FT Peptide 500..505 /label= GXYGX'Y'_motif
FT Peptide 506..511 /label= GXYGX'Y'_motif
FT Peptide 512..517 /label= GXYGX'Y'_motif
FT Peptide 518..523 /label= GXYGX'Y'_motif
FT Peptide 524..529 /label= GXYGX'Y'_motif
FT Peptide 530..535 /label= GXYGX'Y'_motif
FT Peptide 536..541 /label= GXYGX'Y'_motif
FT Peptide 542..547 /label= GXYGX'Y'_motif
FT Peptide 548..553 /label= GXYGX'Y'_motif
FT Peptide 580..585 /label= GXYGX'Y'_motif
FT Peptide 586..591 /label= GXYGX'Y'_motif
FT Peptide 592..597 /label= GXYGX'Y'_motif
FT Peptide 598..603 /label= GXYGX'Y'_motif
FT Peptide 604..609 /label= GXYGX'Y'_motif
FT Peptide 610..615 /label= GXYGX'Y'_motif
FT Peptide 616..621 /label= GXYGX'Y'_motif
FT Peptide 622..627 /label= GXYGX'Y'_motif

```



```
PA (FIBR-) FIBROGEN INC.
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI Clement B, Pihlajaniemi T, Rehn M;
XX
DR WPI; 1999-070292/06.
XX
XX Diagnosis and monitoring of liver disease by measuring collagen type
PT XVII levels - with elevated levels indicative of disease,
PT especially cirrhosis or hepatocellular carcinoma
XX
PS Example 6; Fig 8; 56pp; English.
XX
CC A method has been developed for the detecting liver disease. The method
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific
CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen
CC complex (C) formed as indicator of the amount of Coll18 present; (c)
CC similar analysis of a non-diseased control; and (d) comparing the
CC amounts of Coll18 in the two samples to detect presence or progression of
CC disease. Elevated levels of Coll18 are: (i) indicative of disease,
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,
CC specifically hepatocellular carcinoma (there is a relationship between
CC Coll18 mRNA levels and tumour size and necrosis, and survival times are
CC significantly higher in patients with higher Coll18 levels). The method
CC provides non-invasive, early and accurate diagnosis of liver disease.
CC The present sequence represents the sequence common to mouse alpha-1
CC (XVIII) collagen chain from the present invention.
XX
SQ Sequence 1288 AA;
Query Match 99.7%; Score 920; DB 20; Length 1288;
Best Local Similarity 99.4%; Pred. No. 1.e-100;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQARAVGLSGTFFAFLLSRLLQDIYSI 60
Db 1106 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQARAVGLSGTFFAFLLSRLLQDIYSI 1165
QY 61 VRRADRGSPVIVNLKDEVLSFSDSLFSGSQGLQPGARIFSDGRDVLRRHPAPQKSVW 120
Db 1166 VRRADRGSPVIVNLKDEVLSFSDSLFSGSQGLQPGARIFSDGRDVLRRHPAPQKSVW 1225
QY 121 HGSDPSGRRLMESYCETWRTTGTGQASLLSGLRLLQKAAASCHNSYIVLCIE 175
Db 1226 HGSDPSGRRLMESYCETWRTTGTGQASLLSGLRLLQKAAASCHNSYIVLCIE 1280
Search completed: February 17, 2004, 09:56:33
Job time : 42 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:55:48 ; Search time 21 Seconds

(without alignments)
352,590 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175

Perfect score: 923

Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pcp.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923	100.0	191	4	US-09-561-500-13
2	923	100.0	191	4	US-09-561-108-13
3	923	100.0	191	4	US-09-561-526-13
4	923	100.0	191	4	US-09-561-499-13
5	920	99.7	195	1	US-08-159-784-2
6	874	94.7	185	3	US-08-985-526-36
7	813	88.1	182	4	US-09-561-500-14
8	813	88.1	182	4	US-09-561-108-14
9	813	88.1	182	4	US-09-315-689-3
10	813	88.1	182	4	US-09-561-526-14
11	813	88.1	182	4	US-09-561-499-14
12	813	88.1	183	3	US-09-206-059-2
13	795	86.1	178	4	US-09-315-689-5
14	522	56.6	191	1	US-08-159-784-3
15	150	16.3	35	3	US-09-046-985-2
16	150	16.3	35	3	US-09-474-743-2
17	108	11.7	20	2	US-08-740-168A-1
18	108	11.7	20	3	US-09-349-429-1
19	108	11.7	20	4	US-09-315-689-1
20	108	11.7	20	4	US-09-174-282-1
21	101	10.9	16	3	US-09-385-442-32
22	101	10.9	22	3	US-09-046-985-7
23	101	10.9	22	3	US-09-474-743-7
24	77	8.3	439	4	US-09-252-991A-19623
25	76.5	8.3	1050	4	US-09-428-711A-16
26	76	8.2	505	4	US-09-252-991A-18165
27	75	8.1	190	3	US-09-046-985-15

28	75	8.1	190	3	US-09-474-743-15	Sequence 15, Appl
29	74.5	8.1	577	2	US-08-756-317-13	Sequence 13, Appl
30	74	8.0	6396	4	US-09-410-551B-72	Sequence 72, Appl
31	73.5	8.0	443	4	US-09-252-991A-26460	Sequence 26460, A
32	72.5	7.9	495	4	US-08-679-493A-165	Sequence 165, App
33	72	7.8	347	4	US-09-252-991A-25606	Sequence 25606, A
34	71.5	7.7	1214	2	US-08-231-193A-54	Sequence 54, Appl
35	71.5	7.7	1214	2	US-08-486-273A-54	Sequence 54, Appl
36	71.5	7.7	1214	3	US-08-480-474-54	Sequence 54, Appl
37	71.5	7.7	1214	3	US-08-940-086A-54	Sequence 54, Appl
38	71.5	7.7	1214	4	US-08-935-105A-54	Sequence 54, Appl
39	71.5	7.7	1214	4	US-09-648-797-54	Sequence 54, Appl
40	71.5	7.7	1214	4	US-09-386-123-54	Sequence 54, Appl
41	71.5	7.7	1219	2	US-08-231-193A-50	Sequence 50, Appl
42	71.5	7.7	1219	2	US-08-486-273A-50	Sequence 50, Appl
43	71.5	7.7	1219	3	US-08-480-474-50	Sequence 50, Appl
44	71.5	7.7	1219	3	US-08-940-086A-50	Sequence 50, Appl
45	71.5	7.7	1219	3	US-08-940-086A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1

US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Philip A. Brekken

; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002500

; CURRENT APPLICATION NUMBER: US/09/561,500

; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 60/131,432

; PRIOR FILING DATE: 1999-04-28

; NUMBER OF SEQ ID NOS: 44

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 191

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC

US-09-561-500-13

Query Match 100.0%; Score 923; DB 4; Length 191;

Best Local Similarity 100.0%; Pred. No. 1.9e-106;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOARAVGLSGTFRAFLSRLQDLYSI 60

DB 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOARAVGLSGTFRAFLSRLQDLYSI 67

QY 61 VRRADRGVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRHDPANPKSVW 120

DB 68 VRRADRGVPIVNLKDEVLSFSGSQQLQPGARIFSGDGRDVLRHDPANPKSVW 127

QY 121 HGSDFSGRLMESYCTWRTTGTGATGQASSLSGRLLLEQKAASCHNSYIVLCIE 175

DB 128 HGSDFSGRLMESYCTWRTTGTGATGQASSLSGRLLLEQKAASCHNSYIVLCIE 182

RESULT 2

US-09-561-108-13

; Sequence 13, Application US/09561108

; Patent No. 6342221

; GENERAL INFORMATION:

; APPLICANT: Philip E. Thorpe

; APPLICANT: Philip A. Brekken

; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF

; FILE REFERENCE: 4001.002584

; CURRENT APPLICATION NUMBER: US/09/561,108

```
/ CURRENT FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/131,432
/ PRIOR FILING DATE: 1999-04-28
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 13
/ LENGTH: 191
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-13

Query Match 100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 67

Qy 61 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 120
Db 68 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 127

Qy 121 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 182

RESULT 3
US-09-561-526-13
/ Sequence 13, Application US/09561526
/ Patent No. 6416758
/ GENERAL INFORMATION:
/ APPLICANT: Philip E. Thorpe
/ TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
/ FILE REFERENCE: 4001.002586
/ CURRENT APPLICATION NUMBER: US/09/561,526
/ PRIOR FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/131,432
/ PRIOR FILING DATE: 1999-04-28
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 13
/ LENGTH: 191
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-526-13

Query Match 100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 67

Qy 61 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 120
Db 68 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 127

Qy 121 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 182

RESULT 4
US-09-561-499-13
/ Sequence 13, Application US/09561499
```

```
/ Patent No. 6524583
/ GENERAL INFORMATION:
/ APPLICANT: Philip E. Thorpe
/ APPLICANT: Rolf A. Brekken
/ TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
/ FILE REFERENCE: 4001.002582
/ CURRENT APPLICATION NUMBER: US/09/561,499
/ CURRENT FILING DATE: 2000-04-28
/ PRIOR APPLICATION NUMBER: 60/131,432
/ PRIOR FILING DATE: 1999-04-28
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 13
/ LENGTH: 191
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-499-13

Query Match 100.0%; Score 923; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.9e-106; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 67

Qy 61 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 120
Db 68 VRRADRGSVPIVNLKDEVLSFSGQGLQPGARIFSDGRDVLHRHPAWPKSVW 127

Qy 121 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 175
Db 128 HGSDPSGRRLMESYCYETWRTTGTGATGQASSLLSGLLEQKAASCHNSYIVLCIE 182

RESULT 5
US-08-159-784-2
/ Sequence 2, Application US/08159784
/ Patent No. 5643783
/ GENERAL INFORMATION:
/ APPLICANT: Bjorn R. Olsen
/ TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Fish & Richardson
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
/ COMPUTER: IBM PS/2 Model 50Z or 55SX
/ OPERATING SYSTEM: MS-DOS (Version 5.0)
/ SOFTWARE: WordPerfect (Version 5.1)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/159,784
/ FILING DATE: December 1, 1993
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: John F. Freeman
/ REGISTRATION NUMBER: 29,066
/ REFERENCE/DOCKET NUMBER: 00246/170001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 542-5070
/ TELEFAX: (617) 542-8906
/ TELEX: 200154
/ INFORMATION FOR SEQ ID NO: 2:
```

SEQUENCE CHARACTERISTICS:

LENGTH: 195
TYPE: amino acid
STRANDEDNESS: N/A
TOPOLOGY: N/A
US-08-159-784-2

Query Match 99.7%; Score 920; DB 1; Length 195;
Best Local Similarity 99.4%; Pred. No. 4.6e-106;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 60
DB 12 HTHQDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 71
QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 120
DB 72 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 131
QY 121 HGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 175
DB 132 HGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 186

RESULT 6

US-08-985-526-36
Sequence 36, Application US/08985526
Patent No. 6080728

GENERAL INFORMATION:

APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
FILE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
TITLE OF INVENTION: THERAPY
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: McMorris Jr, Robert G
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613
INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-985-526-36

Query Match 94.7%; Score 874; DB 3; Length 185;
Best Local Similarity 95.5%; Pred. No. 2.2e-100;
Matches 168; Conservative 4; Mismatches 2; Indels 2; Gaps 2;

QY 1 HTHQDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 60
DB 2 HTHQDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 60

QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 119
DB 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 120
QY 120 WHGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 175
DB 121 WHGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 176

RESULT 7

US-09-561-500-14
Sequence 14, Application US/09561500
Patent No. 6342219

GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTI-BODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002500
CURRENT APPLICATION NUMBER: US/09/561,500
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-561-500-14

Query Match 88.1%; Score 813; DB 4; Length 182;

Best Local Similarity 85.7%; Pred. No. 8e-93;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 60
DB 1 HSHRDFQPVHLVALNPLSGMGRGADFCQFQARAVGLSGTFRAPLSSRLQDLYSI 60
QY 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 120
DB 61 VRRADRGSPVIVNLKDEVLSFSGQQLPGARIFSGDRDVLHRHPAPQKSV 120
QY 121 HGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 175
DB 121 HGSDPSGRRLMESYCETWRTTGTGATQASLLSGRLLEQKAASCHNSYIVLCIE 175

RESULT 8

US-09-561-108-14
Sequence 14, Application US/09561108
Patent No. 6342221

GENERAL INFORMATION:

APPLICANT: Philip E. Thorpe
APPLICANT: Rolf A. Brekken
TITLE OF INVENTION: ANTI-BODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
FILE REFERENCE: 4001.002584
CURRENT APPLICATION NUMBER: US/09/561,108
CURRENT FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/131,432
PRIOR FILING DATE: 1999-04-28
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 182
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-561-108-14

```

Query Match      88.1%; Score 813; DB 4; Length 182;
Best Local Similarity 85.7%; Pred. No. 8e-93;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy 1 HTHQDFQVLHVALNTPLSGGMGIRGADPQCFQQARAVLSGTFRAFSSRLQDLYSI 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 HSHRDFQVLHVALNPSLSGGMGIRGADPQCFQQARAVGLAGTFRAFSSRLQDLYSI 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 61 VRRADRGSPVIVNLKDEVLSPWSBLSGSGQQLQPCGARIFSPDGRDVLARHPWPQKSVW 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 VRRADRAAVIVNLKDELLPFSWEALFSGSEGPLKPGARIFSPDCKDVLARHTPWPQKSVW 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 121 HGSDDPSGRRLMESYCETWRTTETTCATQOASLLSGRLLEKAAASCHNSIYVLCIE 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 HGSDDPNGRRLTESYCETWRTPEAPGATQOASLLGRLLGSAASCHHAYIVLCIE 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6348510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-3

```

	Query Match	88.1%;	Score 813;	DB 4;	Length 182;	
	Best Local Similarity	85.7%;	Pred. No. 8e-93;			
	Matches	150;	Conservative 14;	Mismatches 11;	Indels 0;	Gaps 0;
Qy	1	HTHQDFPVLHLVALNTPLSGMRGIRGADFQCFOQARAVCLSTFRAFLSSRLQDIYSI	60			
		: : : : : : : : : : : : : : : : : :				
Dd	1	HSHRDFPVLHLVALNSPLSGMRGIRGADFQCFOQARAVGLACTFRAFLSSRLQDIYSI	60			
		: : : : : : : : : : : : : : : : : :				
Qy	61	VRRADRGSPVIVNLKDELPSWDLPSFGSQQLPGCARIPFSGRDVLHRHPAQKSVM	120			
		: : : : : : : : : : : : : : : : : :				
Dd	61	VRRADRAVPVIVNLKDILLPSWEALPSGEGPKPGCARIPFSPGKDVLRHPTWPQKSVM	120			
		: : : : : : : : : : : : : : : : : :				
Qy	121	HGSDPSGRRLMESYCETWRTEITGATGAASLLSGLLEQAASCHNSYIVLCIE	175			
		: : : : : : : : : : : : : : : : : :				
Dd	121	HGSDPNGRRLTESYCETWRTEAPATGAASLLGRLLGSAACHAYIVLCIE	175			
		: : : : : : : : : : : : : : : : : :				

```

RESULT 10
US-09-561-526-14
; Sequence 14, Application US/09561526
; Patent No. 6416758
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002586
; CURRENT APPLICATION NUMBER: US/09/561,526
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
; OTHER INFORMATION: PEPTIDE  
US-09-561-526-14
```

Query Match 88.1%; Score 813; DB 4; Length 182;
Best Local Similarity 85.7%; Pred.No. 8e-93;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Qy	1	KTHCDPFPVLHLVALNTPLSGGMRGIRGADFCFQQARAVGLSGTFFRAFLSSRLQLDLYSI	60
Dd	1	HSHRDPFVPLHVALNSPLSCMGKRGIGADFCFQQARAVGLAGTFRAFLLSRLQLDLYSI	60
Qy	61	VRRADRGSPIVNLIKDEVLSPSWDSIFPSGCQGLOPGARI FSDPDGRDVLRHPAPWPKSVW	120
Dd	61	VRRADRAAIPVNLIKDELLEFSWEALFSGSEGPIKPGARIFSFDCKDLVRHTPWPKSVW	120
Qy	121	HGSDPSRRRMESYCETWRTETTGTATQAASLLSLGRLLGEAKAAACHSNYSIVLCIE	175
Dd	121	HGSDPNGRILTESYCETWRTEASPATQAASLLCGRLLGSAASHCHAYIVLCIE	175

```
RESULT 11  
US-09-561-499-14  
; Sequence 14, Application US/09561499  
; Patent No. 6524583  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Breken  
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCES: 4001.002582  
; CURRENT APPLICATION NUMBER: US/09/561,499  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
; OTHER INFORMATION: PEPTIDE  
US-09-561-499-14
```

[illegible]

```

RESULT 12
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Nicholas
; APPLICANT: Sim, Kim Lee
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; TITLE OF INVENTION: Proteins and Methods of Use
; FILE REFERENCE: 05213-0370

```

```
; CURRENT APPLICATION NUMBER: US/09/206_059  
; CURRENT FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 183  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-206-059-2
```

Query Match 88.1%; Score 813; DB 3; Length 183;
Best Local Similarity 85.7%; Pred.No.8.1e-93;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY	1	HTHQDFQPVLHLVALNTPLSGGMGRIGADFCQCQAARAVGLSGTTFRAFLSSRLQLDLYSI
DB	1	HSHEDFQPVLLHVALNSPLSGMGRIGADFCCQQAAVAGLACTGFRAFLSSRLQLDLYSI
QY	61	VRRADRGSVPVINLKDEVLPSPWDSLPSGSOQOLPGARIIPSFQGRDVLRHPANPQKSVW
DB	61	VRRADRAAPVIPINUKDELIFFSWALFGSSEGPLKPGAIRIFSPDGKVLRHTPTWPQKSVM
QY	121	HGSDPPSQRRLMESYCETWRTEATTGTATGAQSALLSCRLLEQKAASCHNSIYLCLIE
DB	121	HGSDPNGRRLTESYCETWRTEAPSATGAQSALLGGRLLGSAASHAYIVLCIE

RESULT 13
US-09-315-689-5
Sequence 5, Application US/09315689
Patent No. 6346510
GENERAL INFORMATION:
APPLICANT: O'reilly, Judah
APPlicant: O'Reilly, Michael
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
FILE REFERENCE: US213-0229
CURRENT APPLICATION NUMBER: US/09/315.689
CURRENT FILING DATE: 1999-05-20
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 178
TYPE: PRT
ORGANISM: Homo sapiens
US-09-315-689-5

Query Match 86.1%; Score 795; DB 4; Length 178;
Best Local Similarity 86.5%; Pred.No.1.3e-90;
Matches 148; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY	5	DFQPVHLVALNTPLSGMGRIADPCFOCAARAVGLSGTTFRAFLSSRLQLDIYVERA
DB	1	DFQPVHLVALNSPLSGMGRIADPCFCQAARAVLAGTTFRFAFLSSRLQLDIYVERA
QY	65	DRCGVPIVNLDVLPSPWDSLFSGSOQLOFGARIISFDQRDLVRHPANPQKSVWHGSD
DB	61	DRAAAPVIYNKLDELFIFFSWEALFGSSEGPKFGARIPSFGDKVLRHTPTWPQKSVWHGSD
QY	125	PBGRELMSYCETWRTEATTGTATGAQSALLSCRLLEQKAASCHNSIVILCIE
DB	121	PNGRRLTESYCETWRTEAPSATGAQSALLGGRLLGSAASCHAYIVLCIE

RESULT 14
US-08-159-784-3
Sequence 3, Application US/08159784
Patent No. 5643783
GENERAL INFORMATION:
Applicant: Bjorn R. Olsen
TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson

Tue Feb 17 10:36:40 2004

```

;
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook, David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: CMCC-614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 861-6240
; TELEFAX: (781) 861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-09-046-985-2

Query Match      16.3%; Score 150; DB 3; Length 35;
Best Local Similarity 91.2%; Fred. No. 1.3e-11;
Matches 31; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      36 QARAVGLSGTFRFLSSRLQDIYSIVRRDRGSV 69
      |||||:|||||:|||||:|||||:|||||:
DB      1 QARAVGLAGTFRFLSSRLQDIYSIVRRDRAAV 34

Search completed: February 17, 2004, 09:58:59
Job time : 22 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:56:38 ; Search time 34 Seconds
(without alignments)
1077.703 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175

Perfect score: 923
Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	923	100.0	191	US-09-998-831-13	Sequence 13, Appl
2	923	100.0	191	US-10-373-561-13	Sequence 13, Appl
3	923	100.0	207	US-10-080-797-3	Sequence 3, Appl
4	920	99.7	184	US-10-292-418-18	Sequence 18, Appl
5	920	99.7	184	US-10-131-241-46	Sequence 46, Appl
6	920	99.7	207	US-10-422-934-71	Sequence 71, Appl
7	874	94.7	185	US-10-036-869-36	Sequence 36, Appl
8	813	88.1	180	US-10-131-241-56	Sequence 56, Appl
9	813	88.1	181	US-10-131-241-55	Sequence 55, Appl
10	813	88.1	182	US-09-998-831-14	Sequence 14, Appl
11	813	88.1	182	US-10-373-561-14	Sequence 14, Appl
12	813	88.1	182	US-10-131-241-54	Sequence 54, Appl
13	813	88.1	182	US-10-042-347-3	Sequence 3, Appl
14	813	88.1	183	US-09-873-676-2	Sequence 2, Appl
15	813	88.1	183	US-10-292-418-4	Sequence 4, Appl

16	813	88.1	183	14	US-10-080-797-1	Sequence 1, Appl
17	813	88.1	183	15	US-10-131-241-52	Sequence 52, Appl
18	813	88.1	682	12	US-10-264-049-3010	Sequence 3010, Ap
19	813	88.1	684	11	US-03-961-403-5	Sequence 5, Appl
20	813	88.1	1516	12	US-10-431-642-5	Sequence 3, Appl
21	813	88.1	1516	15	US-10-060-036-166	Sequence 166, App
22	803	87.0	184	12	US-10-292-418-35	Sequence 35, Appl
23	803	87.0	184	15	US-10-131-241-49	Sequence 49, Appl
24	802	86.9	180	15	US-10-131-241-47	Sequence 47, Appl
25	795	86.1	178	15	US-10-131-241-60	Sequence 5, Appl
26	795	86.1	178	15	US-10-042-347-5	Sequence 60, Appl
27	795	86.1	179	15	US-10-131-241-57	Sequence 57, Appl
28	793	85.9	184	12	US-09-938-391-4	Sequence 4, Appl
29	793	85.9	230	12	US-09-938-391-2	Sequence 2, Appl
30	294	31.9	63	9	US-09-822-540A-2	Sequence 2, Appl
31	169	18.3	31	9	US-09-822-540A-1	Sequence 1, Appl
32	108	11.7	20	9	US-09-873-676-115	Sequence 115, App
33	108	11.7	20	10	US-03-405-499-1	Sequence 1, Appl
34	108	11.7	20	10	US-03-174-516-1	Sequence 1, Appl
35	108	11.7	20	10	US-09-154-302-1	Sequence 1, Appl
36	108	11.7	20	12	US-10-351-284-1	Sequence 1, Appl
37	108	11.7	20	15	US-10-131-241-43	Sequence 43, Appl
38	108	11.7	20	15	US-10-232-316-1	Sequence 1, Appl
39	108	11.7	20	15	US-10-042-347-1	Sequence 1, Appl
40	101	10.9	16	10	US-09-766-412-32	Sequence 32, Appl
41	85	9.2	369	12	US-10-029-386-32926	Sequence 32926, A
42	78.5	8.5	403	12	US-10-369-493-3408	Sequence 3408, Ap
43	77	8.3	332	9	US-09-815-242-5659	Sequence 5659, Ap
44	77	8.3	346	9	US-09-815-242-12271	Sequence 12271, A
45	75	8.1	505	12	US-10-369-493-11297	Sequence 11297, A

ALIGNMENTS

RESULT 1
US-09-998-831-13
; Sequence 13, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY
; TITLE OF INVENTION: INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR FILING DATE: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-998-831-13

Query Match	100.0%	Score 923	DB 10	Length 191
Best Local Similarity	100.0%	Pred. No. 2e+96	0	Gaps 0
Matches 175	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	HTHQDFQPVHLVALNTPLSGCMRGIRGADFCQFOQARAVGLSGTFRAPLSRLQDLYSI	60	
Db	8	HTHQDFQPVHLVALNTPLSGCMRGIRGADFCQFOQARAVGLSGTFRAPLSRLQDLYSI	67	
QY	61	VRRADGSPVIVNLKDEVLSFSGSQOLQPOGRIEFGDGRDVLRHPAWPKSVW	120	
Db	68	VRRADGSPVIVNLKDEVLSFSGSQOLQPOGRIEFGDGRDVLRHPAWPKSVW	127	
QY	121	HGSDPSGRRLMESYCTWRTTGTATGQASSLLSRLLEQKAASCHNSYIVLCIE	175	

```
Db 128 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 182
RESULT 2
US-10-373-561-13
; Sequence 13, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-10-373-561-13
Query Match 100.0%; Score 923; DB 12; Length 191;
Best Local Similarity 100.0%; Pred. No. 2e-96;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 8 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 67
QY 61 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 120
Db 68 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 127
QY 121 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 128 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 182
RESULT 3
US-10-080-797-3
; Sequence 3, Application US/10080797
; Publication No. US20020183253A1
; GENERAL INFORMATION:
; APPLICANT: Campochiaro, Peter A.
; APPLICANT: Dixon, Katharine H.
; TITLE OF INVENTION: METHOD FOR TREATING OCULAR
; TITLE OF INVENTION: NEOVASCULARIZATION
; FILE REFERENCE: 4-31881A
; CURRENT APPLICATION NUMBER: US/10/080,797
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Mouse
US-10-080-797-3
Query Match 100.0%; Score 923; DB 14; Length 207;
Best Local Similarity 100.0%; Pred. No. 2.3e-96;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 24 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 83
```

```
QY 61 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 120
Db 84 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 143
QY 121 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 144 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 198
RESULT 4
US-10-292-418-18
; Sequence 18, Application US/10292418
; Publication No. US20030139365A1
; GENERAL INFORMATION:
; APPLICANT: Lo, Kin-Ming
; APPLICANT: Li, Yue
; APPLICANT: Gallies, Stephen D
; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as
; TITLE OF INVENTION: Immunofusins
; FILE REFERENCE: LEX-006C1
; CURRENT APPLICATION NUMBER: US/10/292,418
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 09/383,315
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 60/097,883
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-292-418-18
Query Match 99.7%; Score 920; DB 12; Length 184;
Best Local Similarity 99.4%; Pred. No. 4.2e-96;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
Db 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCFQQAARAVGLSGTFRFLSSRLQDLYSI 60
QY 61 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 120
Db 61 VRRADRGSPVIVNLKDEVLSFSGQSGQLQPGARIFSDGDRDVLHRHPAWPKSVW 120
QY 121 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
Db 121 HGS DSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
RESULT 5
US-10-131-241-46
; Sequence 46, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
```

LENGTH: 184
TYPE: PRT
ORGANISM: Murinae sp.
US-10-131-241-46

Query Match 99.7%; Score 920; DB 15; Length 184;
Best Local Similarity 99.4%; Pred. No. 4.2e-96;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
DB 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
QY 61 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 120
DB 61 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 120

QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 175
DB 121 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 175

RESULT 6
US-10-422-934-71
Sequence 71, Application US/10422934
Publication No. US20030186841A1
GENERAL INFORMATION:
APPLICANT: Barbas, Carlos F., III
APPLICANT: Kadan, Michael
APPLICANT: Beerli, Roger
TITLE OF INVENTION: LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS
FILE REFERENCE: 22908-1227C
CURRENT APPLICATION NUMBER: US/10/422,934
CURRENT FILING DATE: 2003-04-23
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 09/586,625
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: 09/433,042
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 71
LENGTH: 207
TYPE: PRT
ORGANISM: Muridae
US-10-422-934-71

Query Match 99.7%; Score 920; DB 12; Length 207;
Best Local Similarity 99.4%; Pred. No. 4.9e-96;
Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
DB 24 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 83
QY 61 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 120
DB 84 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 143
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 175
DB 144 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 198

RESULT 7
US-10-036-869-36
Sequence 36, Application US/10036869
Publication No. US200201516A1
GENERAL INFORMATION:
APPLICANT: Mixson, James A
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
THERAPY
NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:
ADDRESSEE: Connolly, Bove, Lodge, & Hutz
STREET: 1220 Market Street, P.O. Box 2207
CITY: Wilmington
STATE: Delaware
COUNTRY: U.S.A.
ZIP: 19899

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/036,869
FILING DATE: 29-NO. US200201516A1-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/985,526
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/608,845
FILING DATE: 16-JUL-1996

ATTORNEY/AGENT INFORMATION:
NAME: McMorrow Jr., Robert G
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 658-9141
TELEFAX: (302) 658-5613

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:

LENGTH: 185 amino acids
TYPE: amino acid
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-036-869-36

Query Match 94.7%; Score 874; DB 14; Length 185;
Best Local Similarity 95.5%; Pred. No. 7e-91; 2; Gaps 2;
Matches 168; Conservative 4; Mismatches 2; Indels 2;
QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
DB 2 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
QY 61 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 119
DB 61 VRRADRGSPVIVNLKDEVLSFGSGQGLQPCARIFSDGRDVLHRHPAPQKSV 120
QY 120 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 175
DB 121 HGSDPSGRRLMESYCETWRTTGTATGQASLLSGLRLEQKAASCHNSYIVLCIE 176

RESULT 8
US-10-131-241-56
Sequence 56, Application US/10131241
Publication No. US20030012792A1
GENERAL INFORMATION:
APPLICANT: Holaday, John W.
APPLICANT: Fortier, Anne H.
TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation
FILE REFERENCE: 05213-0344 43170-271565
CURRENT APPLICATION NUMBER: US/10/131,241
CURRENT FILING DATE: 2002-07-22

PRIOR APPLICATION NUMBER: US 09/413,049
PRIOR FILING DATE: 1999-10-06
PRIOR APPLICATION NUMBER: US 09/316,802
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: US 60/086,586
PRIOR FILING DATE: 1998-05-22
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn version 3.1
SEQ ID NO 56
LENGTH: 180

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-56

Query Match      88.1%; Score 813; DB 15; Length 180;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60
DB 1 HSHRDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGDRVLRHPAPQKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSVW 120

QY 121 HGSDPSGRRLMESYCEWTWRTTGTATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175
DB 121 HGSDPNGRRLTESYCEWTWRTTGTATGQASSLLGRLLGQAASCHHAYIVLCIE 175

RESULT 9
US-10-131-241-55
; Sequence 55, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation
; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131.241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 55
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-241-55

Query Match      88.1%; Score 813; DB 15; Length 181;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60
DB 1 HSHRDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGDRVLRHPAPQKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSVW 120

QY 121 HGSDPSGRRLMESYCEWTWRTTGTATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175
DB 121 HGSDPNGRRLTESYCEWTWRTTGTATGQASSLLGRLLGQAASCHHAYIVLCIE 175

RESULT 10
US-09-998-831-14
; Sequence 14, Application US/09998831
; Patent No. US20020119153A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
```

```
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/998,831
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 09/561,108
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-09-998-831-14

Query Match      88.1%; Score 813; DB 10; Length 182;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60
DB 1 HSHRDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGDRVLRHPAPQKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSVW 120

QY 121 HGSDPSGRRLMESYCEWTWRTTGTATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175
DB 121 HGSDPNGRRLTESYCEWTWRTTGTATGQASSLLGRLLGQAASCHHAYIVLCIE 175

RESULT 11
US-10-373-561-14
; Sequence 14, Application US/10373561
; Publication No. US20030175276A1
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002582
; CURRENT APPLICATION NUMBER: US/10/373,561
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/561,499
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; OTHER INFORMATION: PEPTIDE
US-10-373-561-14

Query Match      88.1%; Score 813; DB 12; Length 182;
Best Local Similarity 85.7%; Pred. No. 5.7e-84;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60
DB 1 HSHRDFQPVHLVALNTPLSGGMGIRGADPQCQQAQAVGLSGTFRAPFLSSRLQDLYSI 60

QY 61 VRRADRGSVPIVNLKDEVLSFSGSQQLQPGARIFSDGDRVLRHPAPQKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSDGKDVLRHPTWPKSVW 120

QY 121 HGSDPSGRRLMESYCEWTWRTTGTATGQASSLLSGRLLEKQAASCHNSYIVLCIE 175
DB 121 HGSDPNGRRLTESYCEWTWRTTGTATGQASSLLGRLLGQAASCHHAYIVLCIE 175
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:52:13 ; Search time 21 Seconds
(without alignments)
801.406 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175
Perfect score: 923
Sequence: 1 HTHQDPQPVHLVVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: Piri:*
2: Piri2:*
3: Piri3:*
4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	923	100.0	1774	2 B56101	collagen alpha 1(X
2	920	99.7	1315	2 A56101	collagen alpha 1(X
3	813	88.1	684	2 A53019	collagen alpha 1(X
4	535	58.0	1388	2 A53317	collagen alpha 1(X
5	370	40.1	650	2 T22002	hypothetical prote
6	85	9.2	995	2 H59432	RhodAP protein hom
7	79.5	8.6	286	2 E97241	hypothetical prote
8	79	8.6	427	2 A53798	58K membrane-assoc
9	78.5	8.5	416	2 AD0921	probable lipopolys
10	78	8.5	1055	2 S53597	chlorophyll a/b-bi
11	77.5	8.4	474	2 T10011	probable penicilli
12	77.5	8.4	492	2 B86911	probable penicilli
13	77.5	8.4	6420	2 T30283	polyketide synthas
14	77	8.3	208	2 T45272	methyltransferase
15	77	8.3	351	2 B89781	conserved N-acety
16	77	8.3	534	2 JC7912	amino-acid N-acety
17	77	8.3	1575	2 S68448	synaptotjanin, 170K
18	76.5	8.3	608	2 A97685	hypothetical prote
19	76.5	8.3	608	2 AC2910	hypothetical prote
20	76.5	8.3	996	2 G87687	hypothetical prote
21	76.5	8.3	1054	2 T30933	chitinase (EC 3.2.
22	76	8.2	309	1 E65112	hypothetical 34.6
23	76	8.2	309	2 B91140	hypothetical prote
24	76	8.2	309	2 E85985	hypothetical prote
25	75.5	8.2	4924	2 T50176	probable peptidase
26	74.5	8.1	183	4 S59318	hypothetical prote
27	74.5	8.1	3670	2 T36249	CDA peptidase synthe
28	74.5	8.1	7576	2 T17428	FK506 polyketide s
29	74	8.0	244	2 C45729	sulfolipid biosynt

30 73.5 8.0 488 2 A87569 peptidase M20/M25/
31 73 7.9 477 2 T05202 pectinesterase hom
32 72.5 7.9 353 2 A38131 mocA protein [impo
33 72.5 7.9 353 2 A31156 oxidoreductase moc
34 72.5 7.9 419 2 D42726 nitrite hydratase
35 72.5 7.9 463 2 B59994 xaa-His dipeptidas
36 72.5 7.9 739 1 VHIWEH nucleocapsid prote
37 72.5 7.9 800 2 T25140 nucleocapsid prote
38 72 7.8 309 2 AH0906 conserved hypothet
39 72 7.8 396 2 H95935 probable aminotran
40 72 7.8 442 2 I47074 gene CD5 protein -
41 72 7.8 732 2 C84487 hypothetical prote
42 72 7.8 1955 1 ASCH agrin precursor
43 72 7.8 2899 2 T21546 hypothetical prote
44 71.5 7.7 204 2 B45022 CRK-I - human
45 71.5 7.7 732 2 AF0439 probable Rhs acces

ALIGNMENTS

RESULT 1

B56101
collagen alpha 1(XVIII) chain precursor, long splice form - mouse
N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin
C:Species: Mus musculus (house mouse)
C:Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 31-Mar-2000
C:Accession: B56101; C56101; S72450; S65595; PNO675; A54072; A58816
R:Rehn, M.; Pihlajaniemi, T.
J. Biol. Chem. 270, 4705-4711, 1995

A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
tif homologous to rat and Drosophila frizzled proteins.
A:Reference number: A56101; MUID:95181468; PMID:7876242
A:Accession: B56101

A:Molecule type: mRNA
A:Residues: 1-562 <REH1>

A:CROSS-references: GB:U11637; NID:G618429; PIDN:AAC52179.1; PID:G618430
A:Experimental source: splice form clone PE17.24
A:Accession: C56101

A:Molecule type: mRNA
A:Residues: 1-239,487-562 <REH2>

A:CROSS-references: GB:U11637; NID:G618429
A:Experimental source: splice form clones PE8.1, PE19, PE15.2

R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
submitted to the EMBL Data Library, August 1993

A:Reference number: S72450
A:Accession: S72450

A:Molecule type: mRNA
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16

A:CROSS-references: EMBL:L22545; NID:G348968; PIDN:AA19787.1; PID:G511298
R:Oh, S.P.; Kanagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994

A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
A:Reference number: A58370; MUID:94240111; PMID:8183893

A:Accession: S65595
A:Molecule type: mRNA

A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>
A:CROSS-references: EMBL:L22545

R:Abe, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.
Biochem. Biophys. Res. Commun. 196, 576-582, 1993

A:Title: Identification of a novel collagen chain represented by extensive interruptions
A:Reference number: PNO675; MUID:94059075; PMID:8240330

A:Accession: PNO675
A:Molecule type: mRNA

A:Residues: 635-1774 <ABE>
R:Rehn, M.; Hintikka, E.; Pihlajaniemi, T.

J. Biol. Chem. 269, 13929-13935, 1994

A:Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str
collagen chain.

A:Reference number: A54072; MUID:94245707; PMID:8188673
A:Accession: A54072

A:Molecule type: DNA; mRNA
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>

A;Cross-references: GB:U03714; NID:9487733; PIDN:AAA20657.1; PID:9487734
 R;O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; Bli
 Cell 88, 277-285, 1997
 A;Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.
 A;Reference number: A58816; MUID:97160848; PMID:9008168
 A;Accession: A58816
 A;Molecule type: protein
 A;Residues: 1591-1610 <ORE>
 A;Experimental source: hamangioendothelium cells
 A;Note: inhibits endothelial cell proliferation
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 lated and subsequently O-glycosylated.
 C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
 C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
 ay be useful in treating solid tumors.
 C;Genetics:
 A;Gene: MGI:Coll18a1
 A;Cross-references: MGI:711175
 A;Map position: 10:41.0
 A;Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/
 A;Note: the list of introns is incomplete
 C;Superfamily: unassigned collagens
 C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
 F;1-1774/Product: collagen alpha 1(XVII) chain precursor, long splice form #status pred
 F;1-239,487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form #
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;361-486/Region: frizzled similarity
 F;786-812/Domain: collagenous #status predicted <CO01>
 F;823-896/Domain: collagenous #status predicted <CO02>
 F;921-1042/Domain: collagenous #status predicted <CO03>
 F;1066-1148/Domain: collagenous #status predicted <CO04>
 F;1163-1204/Domain: collagenous #status predicted <CO05>
 F;1218-1290/Domain: collagenous #status predicted <CO06>
 F;1301-1333/Domain: collagenous #status predicted <CO07>
 F;1346-1369/Domain: collagenous #status predicted <CO08>
 F;1351-1353/Region: cell attachment (R-G-D) motif
 F;1377-1428/Domain: collagenous #status predicted <CO09>
 F;1442-1459/Domain: collagenous #status predicted <CO10>
 F;1591-1774/Product: endostatin #status predicted <EST>
 F;1598-1774/Region: multiplexin collagen carboxyl-terminal similarity
 F;354,361,947/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F;910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 923; DB 2; Length 1774;
 Best Local Similarity 100.0%; Pred. No. 6.2e-82;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADFCQFQARAVGLSGTFFAFLLSRLODLYSI 60
 DB 1591 HTHQDFQPVHLVALNTPLSGGMGIRGADFCQFQARAVGLSGTFFAFLLSRLODLYSI 1650

QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSGQQLQPGARIFSPDGRDVLHPAPQKSVW 120
 DB 1651 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSGQQLQPGARIFSPDGRDVLHPAPQKSVW 1710

QY 121 HGSDPSGRRLMESYCTWRTTGTATGQASSLLSGRLLEQKAACHNSYIVLCIE 175
 DB 1711 HGSDPSGRRLMESYCTWRTTGTATGQASSLLSGRLLEQKAACHNSYIVLCIE 1765

RESULT 2
 A56101
 collagen alpha 1(XVIII) chain precursor, short splice form - mouse
 N;Contains: endostatin
 C;Species: Mus musculus (house mouse)
 C;Date: 03-Oct-1995 #sequence_revision 08-May-1998 #text_change 31-Mar-2000
 C;Accession: A56101; A58371; S72450; S65595
 R;Rehn, M.; Pihlajaniemi, T.
 J. Biol. Chem. 270, 4705-4711, 1995
 A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissu
 tif homologous to rat and Drosophila frizzled proteins.
 A;Reference number: A56101; MUID:95181468; PMID:7876242

A;Accession: A56101
 A;Molecule type: mRNA
 A;Residues: 1-103 <REHI>
 A;Cross-references: GB:U11636; NID:9618427; PIDN:AAC52178.1; PID:9618428
 R;Rehn, M.; Pihlajaniemi, T.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994
 A;Title: Alpha1(XVIII), a collagen chain with frequent interruptions in the collagenous
 A;Reference number: A58371; MUID:94240112; PMID:8183894
 A;Accession: A58371
 A;Molecule type: mRNA
 A;Residues: 1-928 <REH2>
 A;Cross-references: GB:L16898; NID:9404754; PIDN:AAA37434.1; PID:9553894
 R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.
 Submitted to the EMBL Data Library, August 1993
 A;Reference number: S72450
 A;Accession: S72450
 A;Molecule type: mRNA
 A;Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHV>
 A;Cross-references: EMBL:L22545; NID:9348968; PIDN:AAA19787.1; PID:9511298
 R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994
 A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa
 A;Reference number: A58370; MUID:94240111; PMID:8183893
 A;Accession: S65595
 A;Molecule type: mRNA
 A;Residues: 28-1315 <OHS>
 A;Cross-references: EMBL:L22545
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
 lated and subsequently O-glycosylated.
 C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri
 C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un
 ay be useful in treating solid tumors.
 C;Genetics:
 A;Gene: MGI:Coll18a1
 A;Cross-references: MGI:711175
 A;Map position: 10:41.0
 C;Superfamily: unassigned collagens
 C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc
 F;1-25/Domain: signal sequence #status predicted <SIG>
 F;24-235/Region: thrombospondin amino-terminal similarity
 F;26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M
 F;327-353/Domain: collagenous #status predicted <CO1>
 F;364-437/Domain: collagenous #status predicted <CO2>
 F;462-583/Domain: collagenous #status predicted <CO3>
 F;607-689/Domain: collagenous #status predicted <CO4>
 F;704-745/Domain: collagenous #status predicted <CO5>
 F;759-831/Domain: collagenous #status predicted <CO6>
 F;842-874/Domain: collagenous #status predicted <CO7>
 F;887-910/Domain: collagenous #status predicted <CO8>
 F;892-894/Region: cell attachment (R-G-D) motif
 F;918-969/Domain: collagenous #status predicted <CO9>
 F;983-1000/Domain: collagenous #status predicted <CO10>
 F;1132-1315/Product: endostatin #status predicted <EST>
 F;1139-1315/Region: multiplexin collagen carboxyl-terminal similarity
 F;126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;172-228/Diulfide bonds: #status predicted
 F;240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F;451,454,1257/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 99.7%; Score 920; DB 2; Length 1315;
 Best Local Similarity 99.4%; Pred. No. 8.5e-82;
 Matches 174; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMGIRGADFCQFQARAVGLSGTFFAFLLSRLODLYSI 60
 DB 1132 HTHQDFQPVHLVALNTPLSGGMGIRGADFCQFQARAVGLSGTFFAFLLSRLODLYSI 1191

QY 61 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSGQQLQPGARIFSPDGRDVLHPAPQKSVW 120
 DB 1192 VRRADRGSPVIVNLKDEVLSPSWDSLFSGSGQQLQPGARIFSPDGRDVLHPAPQKSVW 1251

QY 121 HGSDPSGRRLMESYCTWRTTGTATGQASSLLSGRLLEQKAACHNSYIVLCIE 175

Db 1252 HGSDFSGRRLLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 1306

RESULT 3

A53019 collagen alpha 1(XVII) chain - human (fragment)

N;Contains: endostatin

C;Species: Homo sapiens (man)

C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 31-Mar-2000

C;Accession: A53019

R;Oh, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen, Genomics 19, 494-499, 1994

A;Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization

A;Reference number: A53019; MUID:94245237; PMID:8188291

A;Accession: A53019

A;Molecule type: mRNA

A;Residues: 1-684 <OH>

A;Cross-references: GDB:122548; NID:9348908; PIDN:AAA51864.1; PID:G562794

A;Note: the cited accession number, 125548, is not in Genbank release 103

A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted

C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (labeled and subsequently O-glycosylated).

C;Comment: Different splice forms of collagen alpha 1(XVII) may be involved in perivascular

C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un

ay be useful in treating solid tumors.

C;Genetics:

A;Gene: GDB:COL18A1

A;Cross-references: GDB:138752; OMIM:120328

A;Map position: 21q22.3-21q22.3

C;Superfamily: unassigned collagens

C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc

F;1-684/Product: collagen alpha 1(XVII) chain (fragment) #status predicted <MAT>

F;1-59/Domain: collagenous (fragment) #status predicted <CO4>

F;74-115/Domain: collagenous #status predicted <CO5>

F;129-201/Domain: collagenous #status predicted <CO6>

F;212-244/Domain: collagenous #status predicted <CO7>

F;257-278/Domain: collagenous #status predicted <CO8>

F;262-264/Region: cell attachment (R-G-D) motif

F;286-340/Domain: collagenous #status predicted <CO9>

F;354-371/Domain: collagenous #status predicted <CO10>

F;502-684/Product: endostatin #status predicted <EST>

F;509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 88.1%; Score 813; DB 2; Length 684;

Best Local Similarity 85.7%; Pred. No. 1.2e-71;

Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTQDFQPVHLVALNTPLSGGMRGIRGADFCFQQAARAVLSGTFRAFLLSRQLDLYSI 60

Db 502 HSRDFQPVHLVALNTPLSGGMRGIRGADFCFQQAARAVLSGTFRAFLLSRQLDLYSI 561

QY 61 VRRADRGSPVIVNLKDEVLSPSWSLFSGSGQLOQPCARIFSGDGRDLRHPAPQKSVW 120

Db 562 VRRADRAAVPVLNKLDELFPSEALFSSGGLPKGARIFSGDKDLRHTWFOKSVW 621

QY 121 HGSDFSGRRLLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175

Db 622 HGSDFNPNRLTESYCETWRTTGTATGQASSLLGRLGQSAASCHHAYIVLCIE 676

RESULT 4

A53317 collagen alpha 1(XV) chain precursor - human

N;Alternate names: procollagen alpha 1(XV) chain

C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Mar-2000

C;Accession: A53317; A53146; S28778

R;Kivirikko, S.; Heinauanaeki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.

J. Biol. Chem. 269, 4773-4779, 1994

A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intron

A;Reference number: A53317; MUID:94148920; PMID:8106446

A;Accession: A53317

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1388 <KIV>

A;Cross-references: GDB:125280

A;Note: nucleotide sequence and conceptual translation not complete

R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.

J. Biol. Chem. 269, 4042-4046, 1994

A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple

A;Reference number: A53146; MUID:94140817; PMID:8307960

A;Accession: A53146

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-9, 'S', '11-48, 'V', '50-94, 'A', '96-149, 'A', '151-203, 'V', '205-408, 'A', '410-569 <MUR>

A;Cross-references: GDB:121230; NID:9415605; PIDN:BA04762.1; PID:1005294; PID:9460703

R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.

Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992

A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), charac

A;Reference number: S28778; MUID:93066196; PMID:1279671

A;Accession: S28778

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 544-640, 'P', '642-811, 'P', '813-1252 <MYE>

C;Genetics:

A;Gene: GDB:COL15A1

A;Cross-references: GDB:132578; OMIM:120325

A;Map position: 9q21-9q22

C;Superfamily: unassigned collagens

F;1-22/Domain: signal sequence #status predicted <SIG>

F;23-1388/Product: collagen alpha 1(XV) chain #status predicted <MAT>

F;1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 58.0%; Score 535; DB 2; Length 1388;

Best Local Similarity 60.4%; Pred. No. 5.2e-44;

Matches 102; Conservative 23; Mismatches 40; Indels 4; Gaps 1;

QY 7 QPVHLVALNTPLSGGMRGIRGADFCFQQAARAVLSGTFRAFLLSRQLDLYSIVRRADR 66

Db 1215 KPALHLAALNMPFSGDIR---ADFQCFKQARAAGLLSTYRAFLLSHLQDLSTIVRKAER 1270

QY 67 GSVPTVNLKDEVLSPSWSLFSGSGQLOQPCARIFSGDGRDLRHPAPQKSVWHSQDPS 126

Db 1271 YSLPVLNKGVLFPNNWDSIFSGHGGQFNHPIYSFDGRDIMTPSPQKVIWHGSSPH 1330

QY 127 GRRLMESYCETWRTTGTATGQASSLLSGLLEQKAASCHNSYIVLCIE 175

Db 1331 GVLVDVNYCEAWRTTADTAVTGLASPLSTGKLDQKAYSCANRLIVLCIE 1379

RESULT 5

T22002 hypothetical protein F39H11.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T22002

R;White, S.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19500

A;Accession: T22002

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-650 <WIL>

A;Cross-references: EMBL:Z81079; PIDN:CA803084.1; GSPDB:GN00019; CESP:F39H11.4

A;Experimental source: clone F39H11

C;Genetics:

A;Gene: CESP:F39H11.4

A;Map position: 1

A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 40.1%; Score 370; DB 2; Length 650;

Best Local Similarity 44.4%; Pred. No. 3.2e-28;

Matches 75; Conservative 26; Mismatches 62; Indels 6; Gaps 4;

QY 9 VHLVALNTPLSGGMRGIRGADFCFQQAARAVLSGTFRAFLLSRQLDLYSIVRRADRG 68

Db 464 VIHIALSQPFSGNLHGLRGADLOCYREARAAGYTTTFRAMLSNVQDLVRIHVSDVDFDT 523
QY 69 VPIVNLKDEVLSFGSDSQLOPFGARIFSGDGRDLVRHPAPKQSVHSGSDPSGR 128
Db 524 T-VVNAGHLFPWSRFVNGA--QMPHAKLFSDRDLVNDSDWPDKRVHSGKGGI 580
QY 129 RLMESYCEWTETTTGATGQASSLLSGRLLEQKAAS--CNSYIVLCIE 175
Db 581 R-AEQYCDGMRADSSLTSLAGHSSNTSIFQSSGSEKCNKLVLCVE 628

RESULT 6
H59432
RhoGAP protein homolog [imported] - human
C:Species: Homo sapiens (man)
C>Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C:Accession: H59432
R:Rhodes, S.
submitted to GenBank, October 2001
A:Description: Novel human gene mapping to chromosome 13, similar to rat RhoGAP.
A:Reference number: H59432
A:Accession: H59432
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-995 <RHO>
A:Cross-references: GB:NP_443083; PID:g16445031; PID:NP_443083.1

Query Match 9.2%; Score 85; DB 2; Length 995;
Best Local Similarity 23.4%; Pred. No. 4.7;
Matches 37; Conservative 22; Mismatches 47; Indels 52; Gaps 7;

QY 52 SRLQDLXSVRRDRGSGVP-----IVNLKDEVLSFGSDSQLOPFG 97
Db 36 SRVDVLYLLPRGDRNGSPGTGMTTSSSESLTDLSEPEVCVSIHSESSGSDSRSQPG 95
QY 98 -----ARISFD-----GRDLVRHPAPKQSVHSGSDPSGRRLMESYCEWTET-- 140
Db 96 QCTDNPVMDAPLVSSLPQPRDLVNHPPHK-----NEKTRAPAKSFLKRMETLR 149
QY 141 -----ETTATGATGQASSLLSGRLLEQ-----KAASC 165
Db 150 KGAGHRHKGSGRTG--GLVSGPMLQEPSPFKAMQC 185

RESULT 7
E97241
hypothetical protein CAC2776 [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: E97241
R:Noelling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E97241
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <CUR>
A:Cross-references: GB:AB001437; PID:AAK80720.1; PID:g15025814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2776

Query Match 8.6%; Score 79.5; DB 2; Length 286;
Best Local Similarity 24.1%; Pred. No. 3.6;
Matches 38; Conservative 21; Mismatches 42; Indels 57; Gaps 8;

QY 24 RGIRGADPQQQARAVGLSGTFRFLS-----SRLQDL--YSIVRR-----A 64
Db 42 RGIUVDFISSPIRESI--LSATLKIYSVYNSQLKIKFYSLQKYSINTVYNTQPII 100
QY 65 DRGSPVPIVNLKDEV-----LSPSW-----DSLFSGSGQLQPGARI 100

Db 101 DTNYGVTMTNEINFINVLDSNITSQWHGSGVANLGVMIHGDLSGSGS-----I 151
QY 101 RFDGGRDLVRHPAPKQSVHSGSDPSGRRL-----MESY 134
Db 152 VGFAGISAINSSLRLNVLNVQFSANSGRVLTIYTMESY 189

RESULT 8
A53798
58K membrane-associated protein - rat
N:Alternate names: 58K microfilament-associated protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 24-Nov-1999
C:Accession: A53798
R:Jiang, S.H.; Huang, J.; Li, Y.; Salas, P.J.I.; Fregien, N.; Carraway, C.A.C.; Carraway
J. Biol. Chem. 269, 15067-15075, 1994
A:Title: Molecular cloning and sequencing of a 58-kDa membrane- and microfilament-associ
A:Reference number: A53798; MUID:94253065; PMID:8195143
A:Accession: A53798
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-427 <JUA>
A:Cross-references: GB:U15425; GB:S70214; NID:g562083; PID:AA854085.1; PID:g562084
C:Superfamily: mammalian retrovirus gag polyprotein I
C:Keywords: actin binding; monomer; phosphoprotein

Query Match 8.6%; Score 79; DB 2; Length 427;
Best Local Similarity 21.7%; Pred. No. 6.6;
Matches 40; Conservative 37; Mismatches 79; Indels 28; Gaps 8;

QY 12 LVALNTPILSGMGRGIR----GADPFCQARAVLGSFTFRAPLSSRLQDLXSVRR---A 64
Db 185 LEAPSPVAGLRGRREVAPDSTSQAPFLRQAGCQMQYWPFSAA---DIYNWKQHNPPF 241
QY 65 DRGSPVPIVNLKDEVL---SPSWD-----SLFSSGQ-----LQPGARIFSDGRDLVR- 110
Db 242 SKDPVALNLIIESVLLTQPTWDDIQLLQALLTSEERQVLLLEARKHVLGNGRPTLP 301
QY 111 ---HPAMP-QKSVHSGSDPSGRRLMESYCEWTETTTGATGQASSLLSGRLLEQKAASCH 166
Db 302 EBIIDAFPLTRPDWDFTTAEGRRHLRYQLLLAGLRGAARRPTNLAQVKQVQVAAETP 361
QY 167 NSYI 170
Db 362 SAFL 365

RESULT 9
AD0921
probable lipopolysaccharide biosynthesis protein STY3629 [imported] - Salmonella enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AD0921
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AD0921
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <PAR>
A:Cross-references: GB:AL513382; PID:CAD09390.1; PID:g16504508; GSPDB:GN00176
C:Genetics:
A:Gene: STY3629

Query Match 8.5%; Score 78.5; DB 2; Length 416;
Best Local Similarity 24.4%; Pred. No. 7.1;
Matches 38; Conservative 15; Mismatches 56; Indels 47; Gaps 7;

QY 3 HQDFOPVLHVAL-----NTPLSGCMRGIRGAD-----FQCQOAR 38
| : : : : :
Db 112 HTDYQGLVRVLVQMGIAWANLLALMKGFDRAGNALSLVGSFLGVAAWLYCY---R 168
| : : : : :
QY 39 AVGLSGTFRAFLSRLQDLVSIERRA---DRSGVPIVNLKDEVLSPSWDSLFSGSQGL 94
| : : : : :
Db 169 VGGVEG---ALLGLALVPALVVPVAGIILIKRGIVPLRYLK-----PSWDNGLAGLSKF 220
| : : : : :
QY 95 QPGARISFD-----GRDVLHRPAWPQKSVWHG 122
| : : : : :
Db 221 TLMALITSVTMPVAYVMNRNQLAAHYSWSDVGIMQG 256
| : : : : :
RESULT 10
S53597
Chlorophyll a/b-binding protein (clone GC18 and others) - Euglena gracilis (var. bacilla
C/Species: Euglena gracilis
A/Variety: var. bacillaris
C/Date: 01-Mar-1996 #sequence_revision 19-Apr-1996 #text_change 09-Sep-1997
C/Accession: S53597; S71160; S71481; S19138
R/Schwartzbach, S.D.
submitted to the EMBL Data Library, July 1994
A/Reference number: S53597
A/Accession: S53597
A/Molecule type: DNA
A/Residues: 1-1055 <SCH>
A/Cross-references: EMBL:X61361; NID:9510387; PID:9510388
A/Experimental source: clone GC18; var. bacillaris
A/Note: this is a revision to the sequence from reference S19138
R/Muchhal, U.S.; Schwartzbach, S.D.
Nucleic Acids Res. 22, 5737-5744, 1994
A/Title: Characterization of the unique intron-exon junctions of Euglena gene(s) encoding
A/Reference number: S53596; MUID:95140640; PMID:7838730
A/Accession: S71160
A/Molecule type: DNA
A/Residues: 977-1055 <MUC1>
A/Cross-references: EMBL:X61361
A/Experimental source: clone GC18; var. bacillaris
A/Accession: S71481
A/Molecule type: mRNA
A/Residues: 977-1055 <MUC2>
A/Cross-references: EMBL:X61361
A/Experimental source: clone CLH09; clone CLH22; var. bacillaris
R/Muchhal, U.S.; Schwartzbach, S.D.
Plant Mol. Biol. 18, 287-299, 1992
A/Title: Characterization of a Euglena gene encoding a polyprotein precursor to the light
A/Reference number: S19138; MUID:92119250; PMID:1731990
A/Accession: S19138
A/Molecule type: DNA
A/Residues: 1-374,376-444, 'A', 446-1043, 'AQQLLF', 1048, 'RFSVTL', 1055 <MUC>
A/Cross-references: EMBL:X61361
A/Note: this sequence has been revised in reference S53597
C/Genetics:
A/Introns: 103/3; 137/3; 375/3; 403/1; 444/2; 598/3; 857/2; 976/3; 1043/3
Query Match 8.5%; Score 78; DB 2; Length 1055;
Best Local Similarity 24.2%; Pred. No. 25;
Matches 37; Conservative 19; Mismatches 63; Indels 34; Gaps 5;
QY 20 SGMRGIRGADFCQFOQA---RAVGLSGTFRAFLSRLQDLYS-----IVRRAD 65
| : : : : :
Db 25 SAGSVGDFGADLDPLYPGGPPDPLGLADPEAFLEKVKVQNGRLAWVAIFGFSVQGL 84
| : : : : :
QY 66 RGSVPVNLKDEVLSPSWDSLFSGSQGLQPGARIFSGDGRDVLHRPAWPQKSVWHGSDP 125
| : : : : :
Db 85 TGKGFVENWVDHLTDPFENNLF-----QLTSGFANLATSGRKAKAAPKSNLSQWYGPD- 138
| : : : : :
QY 126 SGRILMSYCYETWRTTGTG-ATGQASSILSGRL 157
| : : : : :
Db 139 -----RAKWLGLPTGEVPSVLTGEL 158
| : : : : :
RESULT 11
T10011
probable penicillin-binding protein - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C/Accession: T10011
R/Cole, S.T.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z16916
A/Accession: T10011
A/Status: preliminary; translated from GB/EMBL/DBSJ
A/Molecule type: DNA
A/Residues: 1-474 <COL>
A/Cross-references: EMBL:Z70722; NID:e1059634; PID:e338505
C/Genetics:
A/Gene: pbpa
Query Match 8.4%; Score 77.5; DB 2; Length 474;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 35; Conservative 18; Mismatches 68; Indels 25; Gaps 5;
QY 35 QQARAVGLSGTFRAFLSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGL 94
| : : : : :
Db 101 EQAEDALLNGSDERLFGRLADFFI--GRDPRGNNVDITINPRVQQTGWDAMQCGGSP 158
| : : : : :
QY 95 QPGARIF-----SFDGRDVLHRPAWPQKSVWH--GSDPS----GRRLMES 133
| : : : : :
Db 159 CKGAVVALEPSTGKILAMVSTPSYDPLNLLASHNPPEQAQAWRLHDDPNPLINRAIS 218
| : : : : :
QY 134 Y--CETWRTTGTGATGQASSILSGRL 157
| : : : : :
Db 219 YPPGSTFKVITTTAALQAGATTSDQL 244
| : : : : :
RESULT 12
B86911
probable penicillin-binding protein [imported] - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C/Accession: B86911
R/Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hc
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A/Title: Massive Gene decay in the leprosy bacillus
A/Reference number: A86909; MUID:21128732; PMID:11234002
A/Accession: B86911
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-492 <STO>
A/Cross-references: GB:AL450380; NID:G13092428; PIDN:CAC29526.1; GSPDB:GNC0147
C/Genetics:
A/Gene: pbpa
Query Match 8.4%; Score 77.5; DB 2; Length 492;
Best Local Similarity 24.0%; Pred. No. 11;
Matches 35; Conservative 18; Mismatches 68; Indels 25; Gaps 5;
QY 35 QQARAVGLSGTFRAFLSRLQDLYSIVRRADRGSVPIVNLKDEVLSPSWDSLFSGSQGL 94
| : : : : :
Db 101 EQAEDALLNGSDERLFGRLADFFI--GRDPRGNNVDITINPRVQQTGWDAMQCGGSP 158
| : : : : :
QY 95 QPGARIF-----SFDGRDVLHRPAWPQKSVWH--GSDPS----GRRLMES 133
| : : : : :
Db 159 CKGAVVALEPSTGKILAMVSTPSYDPLNLLASHNPPEQAQAWRLHDDPNPLINRAIS 218
| : : : : :
QY 134 Y--CETWRTTGTGATGQASSILSGRL 157
| : : : : :
Db 219 YPPGSTFKVITTTAALQAGATTSDQL 244
| : : : : :
RESULT 13

T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C:Species: Streptomyces sp.
A:Variety: strain MA6548
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
R:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A:Title: Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of actinorhodin
A:Reference number: 220806; MUID:97217427; PMID:9063448
A:Accession: T30283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6420 <NOT>
A:Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C:Genetics:
A:Note: fksA
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protein
C:Keywords: carrier protein
F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:3391-3462/Domain: acyl carrier protein homology <ACP1>
F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:3983-4254/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:5307-5378/Domain: acyl carrier protein homology <ACP2>
F:5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:5912-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:6293-6364/Domain: acyl carrier protein homology <ACP3>

Query Match 8.4%; Score 77.5; DB 2; Length 6420;
Best Local Similarity 23.0%; Pred. No. 2.5e+02;
Matches 47; Conservative 25; Mismatches 73; Indels 59; Gaps 9;

QY 4 QDFQVHLHVALNPLSGGMR-----GIRGADFQCFQ-----A 37
DB 5406 RDQEPATVAMACRLPGCVSPRELWLVEGTDATTPAGRGWDLALYDPDPDAVG 5465
QY 38 RAVLSTGTFRAFLSRQLDLYSIYVRADRGSPVIVNLKDEVLPSPWDLFSGGQLQPG 97
DB 5466 KAYSLRGFLGAAEFDAFFDISPESLGMPPQRL---LLETAWAEAI---BRGRINPA 5519
QY 98 ARIFSPGDRVLRHPAMPQKSVHSGDPSPGRRLMESYCEWRTTGTGATGQASLLSGRL 157
DB 5520 ----SLHGREI-----GVYGAAGQYGLGAE-----DTGNATGGSTLLSGRL 5561
QY 158 ----LEOKA-----ASCHNSYIVL 172
DB 5562 AVVLGEGPAVTVTACSSSLVAL 5585

RESULT 14
T45272
methyltransferase homolog [imported] - Streptomyces coelicolor (A3(2))
C:Species: Streptomyces coelicolor
A:Variety: A3(2)
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 11-May-2000
R:Martinez-Costa, O.H.; Martin-Triana, A.J.; Martinez, E.; Fernandez-Moreno, M.A.; Malpa
J. Bacteriol. 181, 4353-4364, 1999
A:Title: An additional regulatory gene for actinorhodin production in Streptomyces livida
A:Reference number: 222953; MUID:99328982; PMID:10400594
A:Accession: T45272
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-208 <MAR>
A:Cross-references: EMBL:Y18817; PIDN:CA851133.1
A:Experimental source: strain J1501
C:Genetics:
A:Note: ORF8

Query Match 8.3%; Score 77; DB 2; Length 208;
Best Local Similarity 25.6%; Pred. No. 4.3;
Matches 40; Conservative 20; Mismatches 52; Indels 44; Gaps 7;

QY 39 AVGLSGTFRFLSSR-----LDLYSIYVRADRGSPV-----IVNLKDEVLPSPWD----- 84
DB 64 AVDISGVAERLAGHARTHGLGDLVDVAVRHLDLRFEPGGRFDLYSA--HYLHTPFDLDR 121
QY 85 SLFSGSQGLQPGARI-----FSFDGRDVLVRHPAMPQKSVHSGDPSPGRRLMESY 134
DB 122 SVLRGAALRPGGRLLVVHGSGTAPWSQDQPDARHPAPQEVADLADFA----- 173
QY 135 CETWTE-----TTGATGQASSLLSGRLLEOKA 162
DB 174 --TWVERAEAPRTATPGGRTAEVWDVLLVRA 207

RESULT 15
B89781
conserved hypothetical protein SA0184 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
R:Kuroda M.; Ohta T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuka, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89759; MUID:21311952; PMID:11418146
A:Accession: B89781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-351 <KUR>
A:Cross-references: GB:BA000018; PID:gl3700106; PIDN:BA841405.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA0184

Query Match 8.3%; Score 77; DB 2; Length 351;
Best Local Similarity 23.7%; Pred. No. 8.1;
Matches 41; Conservative 20; Mismatches 62; Indels 50; Gaps 8;

QY 23 MRGIRGADFQCFQARAVGLSGTFRFLSSRLQDL-----YSIVRRADRG-SVPIV 72
DB 97 IEAIMAAGLKKCLNASIIS-----RELLTSLHQLNDFTLLSFCHNYPRPDTGLSDVLV 151
QY 73 NLKDEVLPSPWDLFSGGQLQPGARIFSPGDRDLR---HPAMPQKSVHSGDP--SG 127
DB 152 NKKNELIY-----QNPRAQIYGFVVGSLRGLPKGLPTTEATRHSHPVVA 199
QY 128 RELMESYCEWRTTGTGATGQASSLLSGRLLEOKAAS-----CHNSYIVLCIE 175
DB 200 KLLQE-----TGVSEVLVGDLSLIEMRQAKQLIDFCRHRHFTLCIE 239

Search completed: February 17, 2004, 09:58:26
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:46:12 ; Search time 17 Seconds

(without alignments)
484.098 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175

Perfect score: 923

Sequence: 1 HTHQDFQVLHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923	100.0	1527	1	CA1H_MOUSE
2	813	88.1	1516	1	CA1H_HUMAN
3	535	58.0	1388	1	CA1E_HUMAN
4	85	9.2	1995	1	SR13_HUMAN
5	77	8.3	1574	1	SYJ1_RAT
6	76.5	8.3	448	1	HGD_BRAJA
7	76.5	8.3	1050	1	ULK1_HUMAN
8	76	8.2	309	1	YHCC_ECOLI
9	76	8.2	1575	1	SVJ1_HUMAN
10	75	8.1	505	1	CATA_METBA
11	75	8.1	1324	1	SVJ1_BOVIN
12	75	8.1	7073	1	RIAB_CVNSA
13	74.5	8.1	611	1	PHBC_RHIME
14	72.5	7.9	419	1	P47K_PSECL
15	72.5	7.9	739	1	VNUC_EBOG4
16	72.5	7.9	800	1	P5CS_CABEL
17	72	7.8	530	1	MATP_MOUSE
18	72	7.8	1233	1	NME3_HUMAN
19	72	7.8	1955	1	AGRI_CHICK
20	71	7.7	1289	1	CSAB_BACUD
21	70.5	7.6	309	1	YEEY_ECOLI
22	70.5	7.6	319	1	HA11_RAT
23	70	7.6	306	1	PERE_ARMRU
24	69.5	7.5	437	1	ADFP_MOUSE
25	69.5	7.5	692	1	GVRB_BABA
26	69.5	7.5	738	1	VNUC_EBOB8
27	69.5	7.5	739	1	VNUC_EBOZ5
28	69.5	7.5	739	1	VNUC_EBOZM
29	69.5	7.5	739	1	VNUC_EBOZM
30	69	7.5	174	1	SZ1A_BRARE
31	69	7.5	174	1	CARQ_MYXXA
32	69	7.5	692	1	VNUC_MABVP
33	69	7.5	695	1	VNUC_MABVP
34	69	7.5	1023	1	DPOL_ADEB3

34	69	7.5	1357	1	YJ03_YEAST	P47104 saccharomyc
35	69	7.5	5217	1	HTS1_COCCA	Q01886 cochliobolu
36	68.5	7.4	266	1	NANH_BACFR	P31206 bacteroides
37	68.5	7.4	1248	1	SYJ2_RAT	O55207 rattus norv
38	68	7.4	362	1	SERC_BACCI	Q59196 bacillus ci
39	68	7.4	587	1	HEMO_HUMAN	P22557 homo sapien
40	68	7.4	735	1	RNSA_MOUSE	Q05921 mus musculu
41	68	7.4	1051	1	ULK1_MOUSE	O70405 mus musculu
42	68	7.4	1203	1	RPA2_YEAST	P21338 saccharomyc
43	68	7.4	1220	1	CSAC_BACUD	P56955 bacillus th
44	68	7.4	1385	1	CSAA_BACUD	Q43760 bacillus th
45	68	7.4	1443	1	SYJ2_HUMAN	O15056 homo sapien

ALIGNMENTS

RESULT 1

ID	CA1H_MOUSE	STANDARD;	PRT; 1527 AA.
AC	P39061; Q61437; Q62002;		
DT	01-FEB-1995 (Rel. 31, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].		
GN	COL18A1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).		
RC	STRAIN=BA1B/C; TISSUE=Liver;		
RX	MEDLINE=94245707; PubMed=8188673;		
RA	Rehn M.V., Hintikka E., Pihlajaniemi T.;		
RT	"Primary structure of the alpha 1 chain of mouse type XVII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVII) chain with its homologue, the alpha 1(XV) collagen chain."		
RL	J. Biol. Chem. 269:13929-13935 (1994).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).		
RA	Rehn M., Hintikka E., Pihlajaniemi T.;		
RT	"Characterization of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters."		
RL	Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE OF 213-1140 FROM N.A. (ISOFORM SHORT).		
RX	MEDLINE=94240112; PubMed=8183894;		
RA	Rehn M.V., Pihlajaniemi T.;		
RT	"Alpha 1(XVII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen."		
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238 (1994).		
RN	[4]		
RP	SEQUENCE OF 240-1527 FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=94240111; PubMed=8183893;		
RA	Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;		
RT	"Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins."		
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233 (1994).		
RN	[5]		
RP	CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.		
RX	MEDLINE=97160848; PubMed=9008168;		
RA	O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;		
RT	"Endostatin: an endogenous inhibitor of angiogenesis and tumor growth."		
RL	Cell 88:277-285 (1997).		

[6]
 RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
 RP MEDLINE=98169382; PubMed=9501087;
 RX Hohenester S., Sasaki T., Olsen B.R., Timpl R.;
 RA "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
 RT resolution.";
 RL EMBL J. 17:1656-1664(1998).
 CC !- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCAN INVOLVED IN GROWTH
 CC FACTOR SIGNALLING.
 CC !- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P39061-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P39061-2; Sequence=VSP 001157, VSP_001158;
 CC !- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC !- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L16898; AAA37434.1; -
 CC EMBL; U03714; AAA20657.1; -
 CC EMBL; U03715; AAC52901.1; JOINED.
 CC EMBL; U34606; AAC52901.1; JOINED.
 CC EMBL; U34608; AAC52901.1; JOINED.
 CC EMBL; U34609; AAC52901.1; JOINED.
 CC EMBL; U34610; AAC52901.1; JOINED.
 CC EMBL; U34611; AAC52901.1; JOINED.
 CC EMBL; U34612; AAC52901.1; JOINED.
 CC EMBL; U34613; AAC52901.1; JOINED.
 CC EMBL; U03716; AAC52901.1; JOINED.
 CC EMBL; U03718; AAC52901.1; JOINED.
 CC EMBL; U03715; AAC52902.1; -
 CC EMBL; U34607; AAC52902.1; JOINED.
 CC EMBL; U34608; AAC52902.1; JOINED.
 CC EMBL; U34609; AAC52902.1; JOINED.
 CC EMBL; U34610; AAC52902.1; JOINED.
 CC EMBL; U34611; AAC52902.1; JOINED.
 CC EMBL; U34612; AAC52902.1; JOINED.
 CC EMBL; U34613; AAC52902.1; JOINED.
 CC EMBL; U03716; AAC52902.1; JOINED.
 CC EMBL; U03718; AAC52902.1; JOINED.
 CC EMBL; U11636; AAC52178.1; -
 CC EMBL; L22545; AAA19787.1; -
 CC PIR; A56101; A56101.
 CC PDB; 1KOE; 16-FEB-99.
 CC PDB; 1DY0; 11-APR-00.
 CC PDB; 1DX1; 21-JAN-01.
 CC MGD; MGI:88451; Coll18a1.
 CC GO; GO:0001525; P:angiogenesis; IMP.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001791; Laminin_G.
 CC InterPro; IPR003129; TSPN.
 CC Pfam; PF01391; Collagen; 8.
 CC Pfam; PF02210; TSPN; 1.
 CC ProDom; PD000007; C1g_helix; 1.
 CC SMART; SM00282; LamG; 1.
 CC SMART; SM00210; TSPN; 1.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 CC 3D-structure.
 CC SIGNAL. 1 26 POTENTIAL.
 CC CHAIN 27 1527 COLLAGEN ALPHA 1 (XVIII) CHAIN.

FT CHAIN 1344 1527
 FT DOMAIN 27 538
 FT DOMAIN 539 565
 FT DOMAIN 566 575
 FT DOMAIN 576 649
 FT DOMAIN 650 673
 FT DOMAIN 674 795
 FT DOMAIN 796 818
 FT DOMAIN 819 901
 FT DOMAIN 902 915
 FT DOMAIN 916 957
 FT DOMAIN 958 970
 FT DOMAIN 971 1043
 FT DOMAIN 1044 1053
 FT DOMAIN 1054 1086
 FT DOMAIN 1087 1098
 FT DOMAIN 1099 1122
 FT DOMAIN 1123 1129
 FT DOMAIN 1130 1181
 FT DOMAIN 1182 1194
 FT DOMAIN 1195 1212
 FT DOMAIN 1213 1527
 FT CARBOHYD 338 338
 FT CARBOHYD 700 700
 FT DISULFID 1376 1516
 FT DISULFID 1478 1508
 FT SITE 1104 1106
 FT VARSPLIC 1 212
 FT VARSPLIC 213 238
 FT CONFLICT 900 900
 FT CONFLICT 947 947
 FT CONFLICT 964 964
 FT CONFLICT 1157 1157
 FT CONFLICT 1266 1266
 FT CONFLICT 1276 1276
 FT CONFLICT 1437 1437
 FT STRAND 1353 1357
 FT STRAND 1362 1362
 FT HELIX 1368 1382
 FT TURN 1383 1384
 FT STRAND 1389 1391
 FT STRAND 1395 1395
 FT TURN 1396 1397
 FT STRAND 1398 1398
 FT HELIX 1400 1403
 FT HELIX 1406 1408
 FT TURN 1409 1411
 FT STRAND 1414 1415
 FT TURN 1417 1418
 FT STRAND 1421 1423
 FT STRAND 1426 1429
 FT TURN 1430 1430
 FT STRAND 1436 1436
 FT TURN 1439 1440
 FT STRAND 1444 1445
 FT TURN 1446 1447
 FT STRAND 1450 1450
 FT TURN 1451 1453
 FT TURN 1455 1455
 FT STRAND 1461 1463
 FT STRAND 1466 1466
 FT TURN 1468 1469
 FT STRAND 1472 1472
 FT TURN 1474 1475
 FT STRAND 1477 1477
 FT HELIX 1478 1481
 FT TURN 1482 1482
 FT TURN 1486 1487
 FT TURN 1489 1494
 FT HELIX 1495 1497

ENDOSTATIN.
 NONHELIICAL REGION 1 (NC1).
 TRIPLE-HELIICAL REGION 2 (NC2).
 TRIPLE-HELIICAL REGION 3 (NC3).
 NONHELIICAL REGION 4 (NC4).
 TRIPLE-HELIICAL REGION 5 (NC5).
 TRIPLE-HELIICAL REGION 6 (NC6).
 TRIPLE-HELIICAL REGION 7 (NC7).
 TRIPLE-HELIICAL REGION 8 (NC8).
 TRIPLE-HELIICAL REGION 9 (NC9).
 TRIPLE-HELIICAL REGION 10 (NC10).
 NONHELIICAL REGION 11 (NC11).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).
 Missing (in isoform Short).
 /FTId=VSP 001157.
 AVPTQLPFPQSNLQAPGRSPAPPDF -> MAPRHHLLDVL
 TSLVLLLVARVSWAE (in isoform Short).
 /FTId=VSP 001158.
 P -> L (IN REF. 4).
 P -> F (IN REF. 4).
 R -> P (IN REF. 4).
 A -> P (IN REF. 4).
 L -> F (IN REF. 4).
 L -> V (IN REF. 4).

Query Match 100.0%; Score 923; DB 1; Length 1527;
 Best Local Similarity 100.0%; Pred. No. 1.3e-81;
 Matches 175; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HTHQDFQVHLVALNTPLSGGMGIRGADFCQCARAVLSTGTFRAFLSSRLQDIYSI 60
 DB 1344 HTHQDFQVHLVALNTPLSGGMGIRGADFCQCARAVLSTGTFRAFLSSRLQDIYSI 1403

OY 61 VRRDRGSPVIVNLKDEVLSPWSLPSGSGQGLQPGARIPSFGRDVLHRPAWPKQSVW 120
 DB 1404 VRRDRGSPVIVNLKDEVLSPWSLPSGSGQGLQPGARIPSFGRDVLHRPAWPKQSVW 1463

OY 121 HGSPPSGRLMESCYETWRTTTCATGQASLLSGRLLEOKAACHNSYIVLCIE 175
 DB 1464 HGSPPSGRLMESCYETWRTTTCATGQASLLSGRLLEOKAACHNSYIVLCIE 1518

RESULT 2
 CALH HUMAN
 ID _CALH HUMAN STANDARD; PRT; 1516 AA.
 AC P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].
 GN COL18A1.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98164096; PubMed=9503365;
 RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;
 RT "Complete primary structure of two variant forms of human type XVII
 collagen and tissue-specific differences in the expression of the
 RT corresponding transcripts.";
 RL Matrix Biol. 16:319-328(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,
 RA Muroshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [3]
 RP SEQUENCE OF 834-1516 FROM N.A.
 RX MEDLINE=94245237; PubMed=8188291;
 RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
 RA Olsen B.R.;
 RT "Cloning of cDNA and genomic DNA encoding human type XVII collagen
 RT and localization of the alpha 1(XVII) collagen gene to mouse
 RT chromosome 10 and human chromosome 21.";
 RL Genomics 19:494-499(1994).
 RN [4]
 RP SEQUENCE OF 1334-1516 FROM N.A.
 RX MEDLINE=94245237; PubMed=8188291;
 RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
 RA Olsen B.R.;
 RT "Cloning of cDNA and genomic DNA encoding human type XVII collagen
 RT and localization of the alpha 1(XVII) collagen gene to mouse
 RT chromosome 10 and human chromosome 21.";
 RL Genomics 19:494-499(1994).
 RN [5]
 RP Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RP INVOLVEMENT IN KNOBLOCH SYNDROME.
 RX MEDLINE=20400145; PubMed=10942434;
 RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
 RA Passos-Bueno M.R.;
 RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
 RT and tumor growth, plays a critical role in the maintenance of retinal
 RT structure and in neural tube closure.";
 RL Hum. Mol. Genet. 9:2051-2058(2000).
 RN [6]
 RP VARIANT ASN-1437.
 RX MEDLINE=21518361; PubMed=11606364;
 RA Lughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
 RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
 RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
 RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
 RT for the development of prostatic adenocarcinoma.";
 RL Cancer Res. 61:7375-7378(2001).
 CC -|- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE
 CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.
 CC -|- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH
 CC FACTOR SIGNALLING.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long; Synonyms=NC-493;
 CC IsoId=P39060-1; Sequence=Displayed;
 CC Name=Short; Synonyms=NC1-303;
 CC IsoId=P39060-2; Sequence=VSP 001155; VSP 001156;
 CC -|- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS
 CC IN LIVER, LUNG AND KIDNEY.
 CC -|- PTM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -|- POLYMORPHISM: There is an association between a polymorphism in
 CC position 1437 and prostate cancer. Heterozygous Asn-1437
 CC individuals have a 2.5 times increased chance of developing
 CC prostate cancer as compared with homozygous Asp-1437 individuals.
 CC -|- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)
 CC [MIM:267750]; an autosomal recessive disorder defined by the
 CC occurrence of high myopia, vitreoretinal degeneration with retinal
 CC detachment, macular abnormalities and occipital encephalocele.
 CC -|- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF018081; AAC39658.1; -;
 CC EMBL; AF018082; AAC39659.1; -;
 CC EMBL; AL163302; CAB90482.1; -;
 CC EMBL; L22548; AAA51864.1; -;
 CC EMBL; AF184060; AAF01310.1; ALT_INIT.
 CC PDB; 1BNL; 02-DEC-98.
 CC GlycosuiteDB; P39060; -;
 CC Genew; HGNC:2195; COL18A1.
 CC MIM; 120328; -;
 CC MIM; 267750; -;
 CC GO; GO:0003581; C:collagen; TAS.
 CC GO; GO:000181; P:tumor suppressor; TAS.
 CC GO; GO:0007397; P:histogenesis and organogenesis; TAS.
 CC GO; GO:0003285; P:negative regulation of cell proliferation; TAS.
 CC GO; GO:0007048; P:oncogenesis; TAS.
 CC GO; GO:0007601; P:vision; TAS.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR001791; Laminin G.
 CC InterPro; IPR003129; TSPN.
 CC Pfam; PF01391; Collagen; 7.
 CC Pfam; PF02110; TSPN; 1.


```

DR MIM; 120325; --
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SV00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Cell adhesion; Collagen; Glycoprotein; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1388
FT DOMAIN 26 555
FT DOMAIN 26 555
FT DOMAIN 556 573
FT DOMAIN 574 618
FT DOMAIN 619 732
FT DOMAIN 733 763
FT DOMAIN 764 798
FT DOMAIN 799 822
FT DOMAIN 823 867
FT DOMAIN 868 878
FT DOMAIN 879 949
FT DOMAIN 950 983
FT DOMAIN 984 1013
FT DOMAIN 1014 1027
FT DOMAIN 1028 1045
FT DOMAIN 1046 1052
FT DOMAIN 1053 1107
FT DOMAIN 1108 1117
FT DOMAIN 1118 1132
FT DOMAIN 1133 1388
FT REPEAT 358 555
FT REPEAT 358 408
FT REPEAT 409 459
FT REPEAT 460 509
FT REPEAT 510 555
FT CARBOHYD 306 306
FT CARBOHYD 324 324
FT CARBOHYD 687 687
FT CARBOHYD 807 807
FT CARBOHYD 814 814
FT CARBOHYD 1046 1046
FT CONFLICT 10 10
FT CONFLICT 49 49
FT CONFLICT 95 95
FT CONFLICT 150 150
FT CONFLICT 204 204
FT CONFLICT 409 409
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;

Query Match 58.0%; Score 535; DB 1; Length 1388;
Best Local Similarity 60.4%; Pred. No. 6.1e-44;
Matches 102; Conservative 23; Mismatches 40; Indels 4; Gaps 1;

QY 7 QPVLHLVALNPTLPGCMRGIRGAPQCFQQAQAVGLSGTFRFLSSRLQDLYSIVRRADR 66
DB 1215 KPAHLAALNPPFSGDIR---ADQCFQQAAGLLSTYFAFLSHLQDLYSIVRKAER 1270

QY 67 GSVPTVNLKDEVLSFSGQGLQOPGARIFSDGRVLRHPAPQKSVHSGDPS 126
DB 1271 YSLPTVNLKGVLFNNWDSIFSGHGQFMNHIPIYSFSDGRDITDPSNPQKSVHSGSPH 1330

QY 127 GRLLMESYCTWRTTTCATGQASLLSGRLLEOKAASCHNSYIVLCIE 175
DB 1331 GVLVDNICEAWRTADTAVTGLASPLSTGKILDKAYSCANRLVLCIE 1379

RESULT 4
SF13 HUMAN
ID SR13 HUMAN STANDARD; PRT; 995 AA.
AC Q9Y3M6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE STAR-related lipid transfer protein 13 (STARL13) (START domain-
DE containing protein 13) (46523.2).
GN STARD13 OR G7850.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Hunt A.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC 1- FUNCTION: May function as a GTPase-activating protein.
CC 1- SIMILARITY: Contains 1 Rho-GAP domain.
CC 1- SIMILARITY: Contains 1 START domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL049801; CAB42562.1; --
CC EMBL; Z84483; CAC94774.1; --
CC PIR; H59432; H59432.
CC Genew; HGNC:13164; STARD13
CC InterPro; IPR000198; RhoGAP.
CC InterPro; IPR002913; START.
CC Pfam; PF00620; RhoGAP; 1.
CC Pfam; PF01852; START; 1.
CC SMART; SM00324; RhoGAP; 1.
CC SMART; SM00234; START; 1.
CC PROSITE; PS00238; RHOAP; 1.
CC PROSITE; PS00948; START; 1.
KW GTPase activation.
FT DOMAIN 545 750 RHO-GAP.
FT DOMAIN 781 989 START.
SQ SEQUENCE 995 AA; 111191 MW; 3F608FA94A4EF8BF CRC64;

Query Match 9.2%; Score 85; DB 1; Length 995;
Best Local Similarity 23.4%; Pred. No. 2.4;
Matches 37; Conservative 22; Mismatches 47; Indels 52; Gaps 7;

QY 52 SRLQDLYSIVRRADRGVSP-----IVNLKDEVLSFSGQGLQOPG 97
DB 36 SRVDDLYTLPRGDRNGSPGGTGMTTSSVLTDLSEPEVCSIHSESSGGSDRSQPG 95

QY 98 -----ARISFD-----GRVLRHPAPQKSVHSGDPSGRLMESYCTWRT-- 140
DB 96 QCOTDNPNVMDAPLVSSSLPQPPRDVNLNHPFPK-----NEKPTRAKSFLKRMETLR 149

QY 141 -----ETTQATGQASLLSGRLLEQ-----KAASC 165
DB 150 GKGAHGRHKGSGRTG--GLVISGPMLOQEPESFKAMQC 185

RESULT 5
SYJ1 RAT
ID SVJ1 RAT STANDARD; PRT; 1574 AA.
AC Q62910; O89092; Q62911;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Synaptojanin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-
DE phosphate 1).
DE SYNJI
GN SYNJI
OS Rattus norvegicus (Rat).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=96149250; PubMed=8552192;
 RA McPherson P.S., Garcia E.P., Slepnev V.I., David C., Zhang X.,
 RA Grabs D., Sosin W.S., Bauerfeind R., Nemoto Y., De Camilli P.;
 RT "A presynaptic inositol-5-phosphatase";
 RL Nature 379:353-357(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Brain;
 RX MEDLINE=98374013; PubMed=9710239;
 RA Wocholowski R., Ranan P.M., Radley E., Parker P.J.;
 RT "Identification and characterization of a novel splice variant of
 RT synaptotagmin1";
 RL FEBS Lett. 432:5-8(1998).
 RN [3]
 RP ALTERNATIVE SPLICING.
 RC TISSUE=Brain;
 RX MEDLINE=96394655; PubMed=8798761;
 RA Ramjaun A.R., McPherson P.S.;
 RT "Tissue-specific alternative splicing generates two synaptotagmin
 RT isoforms with differential membrane binding properties";
 RL J. Biol. Chem. 271:24856-24861(1996).
 CC -!- FUNCTION: Inositol 5-phosphatase which has a role in clathrin-
 CC mediated endocytosis.
 CC -!- CATALYTIC ACTIVITY: 1-phosphatidyl-ID-myo-inositol 4,5-
 CC bisphosphate + H(2)O = 1-phosphatidyl-ID-myo-inositol 4-phosphate
 CC + phosphate.
 CC -!- SUBCELLULAR LOCATION: LOCALIZED MAINLY IN THE SOLUBLE FRACTION (BY
 CC SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=6;
 CC Name=1; Synonyms=170 kDa;
 CC IsoId=Q62910-3; Sequence=Displayed;
 CC Note=A stop codon in position 1309 is suppressed due to
 CC alternative splicing;
 CC Name=2; Synonyms=145 kDa;
 CC IsoId=Q62910-2; Sequence=VSP_002686;
 CC Name=3; Synonyms=Delta-SAC;
 CC IsoId=Q62910-3; Sequence=VSP_002684;
 CC Name=4; Synonyms=170 kDa-16AA;
 CC IsoId=Q62910-4; Sequence=VSP_002685;
 CC Name=5; Synonyms=145 kDa-16AA;
 CC IsoId=Q62910-5; Sequence=VSP_002685; VSP_002686;
 CC Name=6; Synonyms=Delta-SAC-16AA;
 CC IsoId=Q62910-6; Sequence=VSP_002684; VSP_002685;
 CC -!- TISSUE SPECIFICITY: ISOFORM 1 IS FOUND IN NEONATAL BRAIN, AND IN A
 CC WIDE VARIETY OF ADULT NON-NEURONAL TISSUES. ISOFORM 2 IS EXPRESSED
 CC PREDOMINANTLY IN THE NEURONS, BUT IS ALSO FOUND IN ALL OTHER IN THE
 CC TISSUES AT MUCH LOWER LEVELS. ISOFORMS 1 AND 2 ARE DETECTED IN THE
 CC LUNG AND HEART. ISOFORM 1 IS EXPRESSED AT HIGHER LEVELS THAN
 CC ISOFORM 2 IN THE TESTIS AND LIVER AND BOTH ISOFORMS ARE NOT
 CC DETECTED IN THE SKELETAL MUSCLE. ISOFORM 3 WITH THE 16-AMINO-ACID
 CC INSERT IS ONLY FOUND IN THE BRAIN WHILE ISOFORM 3 WITHOUT THE 16-
 CC AMINO-ACID INSERT IS FOUND IN THE LUNG.
 CC -!- DEVELOPMENTAL STAGE: AT EMBRYONIC DAY 12 (E12) ONLY ISOFORM 1 IS
 CC SEEN WHILE AT E16 AND E18 ISOFORMS 1 AND 2 ARE SEEN. IN THE ADULT
 CC BRAIN EXPRESSION OF ISOFORM 2 INCREASES DRAMATICALLY AS COMPARED
 CC WITH ITS EXPRESSION IN EMBRYONIC BRAIN WHERE AS ISOFORM 1
 CC DECREASES TO UNDETECTABLE LEVELS.
 CC -!- DOMAIN: BINDS TO EP15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
 CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS (BY
 CC SIMILARITY).
 CC -!- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,
 CC SH3P4, SH3P8, SH3P13 AND GRB2.
 CC -!- DOMAIN: SPLICING OF THE SAC1 DOMAIN DOES NOT ALTER THE CATALYTIC
 CC ACTIVITY OF SYNAPTOTAGMIN 1.
 CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-

CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
 CC -!- SIMILARITY: Contains 1 SAC domain.
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U45479; AAB60525.1; -;
 CC EMBL; U45479; AAB60526.1; -;
 CC EMBL; AJ006855; CAA07267.1; ALT TERM.
 CC InterPro: IPR005135; Exo_endo_phos.
 CC InterPro: IPR000300; IPPC.
 CC InterPro: IPR000504; RNA_rec_mot.
 CC InterPro: IPR002013; SyJa_N.
 CC Pfam; PF03372; Exo_endo_phos; 1.
 CC Pfam; PF02383; SyJa_N; 1.
 CC SMART; SM00128; IPPC; 1.
 CC PROSITE; P850102; RRM; 1.
 CC PROSITE; P850275; SAC; 1.
 CC Hydrolase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
 KW Multigene family.
 FT DOMAIN 119 442 SAC.
 FT DOMAIN 500 899 CATALYTIC (POTENTIAL).
 FT DOMAIN 894 971 RNA-BINDING (RRM).
 FT DOMAIN 900 1574 PRO-RICH.
 FT DOMAIN 1033 1036 POLY-RICH.
 FT DOMAIN 1105 1110 POLY-PRO.
 FT DOMAIN 1123 1126 POLY-PRO.
 FT DOMAIN 1537 1545 POLY-PRO.
 FT DOMAIN 1401 1423 3 X 3 AA REPEATS OF N-P-P.
 FT REPEAT 1401 1403 1.
 FT REPEAT 1410 1412 2.
 FT REPEAT 1421 1423 3.
 FT VARSPLIC 1 400 Missing (in isoform 3 and isoform 6).
 FT VARSPLIC 1140 1155 /FTID=VSP_002684.
 FT VARSPLIC 1309 1574 Missing (in isoform 4, isoform 5 and
 FT VARSPLIC 588 588 /FTID=VSP_002685.
 FT CONFLICT 1140 1140 G -> D (IN REF. 1; AAB60525).
 FT CONFLICT 1156 1156 G -> GG (IN REF. 2).
 FT CONFLICT 1574 1574 MISSING (IN REF. 2).
 FT SEQUENCE 1574 AA; 172880 MW; 9DEEC876BAB3ADF CRC64;
 Query Match 8.3%; Score 77; DB 1; Length 1574;
 Best Local Similarity 22.6%; Pred. No. 25;
 Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;
 QY 8 PVHLVALNTPLS--GGMGIRGA-----DFQCFOAR-AVGLSGTFAFLSSRLQ 55
 Db 648 PFIRDVAVTVTGMGGATGNKGAIVRMFLHTTSLCFVCSHFAAGSQ-----VKERNE 702
 QY 56 DLYSVIRRAD-----RGSVPINLKDVLSPSWDSLFSGSQ---- 91
 Db 703 DFVEIARKLSFPMGRMLFSDYVFCGDNFYRIDLPNEEVKELIRQONWDSLIAGDQLIN 762
 QY 92 ----GQLQPG-----ARIFS---FGDRVLRHPANPQKSVW 120
 Db 763 QKNAGQIFRGLGKVTFAFTFYKYDLFSEDTSEKCRTPAWTDRVLW 810
 RESULT 6
 HGD_BRAJA STANDARD; PRT; 448 AA.
 ID_HGD_BRAJA
 AC Q89XH1;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Homogentisate 1,2-dioxygenase (EC 1.13.11.5) (Homogentisicase)
DE (Homogentisate oxygenase) (Homogentisic acid oxidase).
GN HMG OR BLL0343.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=USDA 110;
RC MEDLINE=22484998; PubMed=12597275;
RX Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
CC -!- CATALYTIC ACTIVITY: Homogentisate + O(2) = 4-maleylacetate.
CC -!- COFACTOR: Iron (By similarity).
CC -!- PATHWAY: Catabolism of tyrosine; third step.
CC -!- PATHWAY: Catabolism of phenylalanine; fourth step.
CC -!- SIMILARITY: Belongs to the homogentisate dioxygenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005936; BAC45608.1; ALT_INIT.
DR HAMAP; MF 00334; -; 1.
DR Oxidoreductase; Dioxygenase; Metal-binding; Iron;
KW Phenylalanine catabolism; Tyrosine catabolism; Complete proteome.
FT METAL 346 346 IRON (BY SIMILARITY).
FT METAL 352 352 IRON (BY SIMILARITY).
FT METAL 382 382 IRON (BY SIMILARITY).
SQ SEQUENCE 448 AA; 49780 MW; F6DD465868735D3C CRC64;
Query Match 8.3%; Score 76.5; DB 1; Length 448;
Best Local Similarity 25.0%; Pred. No. 6;
Matches 32; Conservative 22; Mismatches 51; Indels 23; Gaps 7;
QY 25 GIRGADFCFOQARAVGLSGTFRA-----FLSRLQDLYSIVRRADRGSVF-IVNLKDE 77
Db 320 GTANIDFVIFPE-RMVADNTFRPPWYHNMIMSEFMGLIYGVYDAKPGQFVPGMSLHNC 378
QY 78 VL--SPSWDSLFSGSQQLQP-----GARIFSPDQDVLRLHPAPKQSVWHSQDPGSRRLM 131
Db 379 MLPHGPRDPAFHASNGELKPVKLTGTMAFMFETR-----YPOVRTAAANAS--TLQ 429
QY 132 BSYCETWR 139
Db 430 DDYADCKW 437
RESULT 7
ULK1_HUMAN
ID ULK1_HUMAN STANDARD; PRT; 1050 AA.
AC 075385;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase ULK1 (EC 2.7.1.-) (Unc-51-like kinase
DE 1).
GN ULK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN SEQUENCE FROM N.A.
RP MEDLINE=98360094; PubMed=9893035;
RA Kuroyanagi H., Yan J., Seki N., Yamanouchi Y., Suzuki Y.-I.,
RA Takaro T., Muramatsu M.-A., Shirasawa T.;
RT "Human ULK1, a novel serine/threonine kinase related to UNC-51 kinase
RT of Caenorhabditis elegans: cDNA cloning, expression, and chromosomal
RT assignment.";
RL Genomics 51:76-85(1998).
[2]
RN INTERACTION WITH GABARAP AND GABARAPL2.
RP MEDLINE=21066693; PubMed=11146101;
RA Okazaki N., Yan J., Yuasa S., Ueno T., Kominami E., Masuho Y.,
RA Koga H., Muramatsu M.-A.;
RT "Interaction of the Unc-51-like kinase and microtubule-associated
RT protein light chain 3 related proteins in the brain: possible role of
RT vesicular transport in axonal elongation.";
RL Brain Res. Mol. Brain Res. 85:1-12(2000).
CC -!- SUBUNIT: Interacts with GABARAP and GABARAPL2.
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed. Detected in the
CC following adult tissues: skeletal muscle, heart, pancreas, brain,
CC placenta, liver, kidney, and lung.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC APGL/UNC-51/ULK1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF045458; AAC32326.1; -;
DR HSSP; P24941; 1CKP.
DR Genew; HGNC:12558; ULK1.
DR MIM; 603168;
CC GO; GO:0004674; P:protein serine/threonine kinase activity; TAS.
CC GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR001245; Tyr kinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 16 278 PROTEIN_KINASE
FT NP_BIND 22 30 ATP (BY SIMILARITY).
FT DOMAIN 287 416 INTERACTION WITH GABARAP AND GABARAPL2.
FT BINDING 46 46 ATP (BY SIMILARITY).
FT ACT_SITE 138 138 BY SIMILARITY.
FT DOMAIN 297 310 POLY-SER.
SQ SEQUENCE 1050 AA; 112601 MW; 4ED9B94028E3C138 CRC64;
Query Match 8.3%; Score 76.5; DB 1; Length 1050;
Best Local Similarity 30.3%; Pred. No. 17;
Matches 43; Conservative 13; Mismatches 47; Indels 39; Gaps 9;
QY 18 PLSGGMRGIRGADFCFOQARAVGLSGTF-RAFLSSRLQDLYSIVRRA-----DRGSVP 70
Db 674 PLGPGGLR--PGED-----PKGPFGRFSFSLRDL--LKAARFGTQAPDPSGTE 718
QY 71 IVNLKDEVLSPWSLFSGSQQLQP-----GARIFSPDQDVLRLHPAPKQSVWHSQDPGSR 129
Db 719 SLQEKPMELAPS-----AGFGSLHPGARAGTSS-----ESPVFTVSGSPSGST 764
QY 130 LMSYCETWRTTGTGATGOASS 151

765 PPQG--PRTRMFSAGPTGSASS 784

```

Db  InterPro: IPR005911; Cons_hypoth1212.
DR  InterPro: IPR006638; ELP3.
DR  Pfam: PF04055; Radical SAM; 1.
DR  SMART: SM00729; ELP3; 1.
DR  TIGRFAMs: TIGR01212; TIGR01212; 1.
KW  Hypothetical protein; Complete proteome.
FT  CONFLICT 24 25   KL -> NV (IN REF. 4).
SQ  SEQUENCE 309 AA; 34606 MW; 61B3187BB77CA19 CRC64;

Query Match      8.2%; Score 76; DB 1; Length 309;
Best Local Similarity 26.1%; Pred. No. 4.3;
Matches 35; Conservative 23; Mismatches 42; Indels 34; Gaps 10;

QY  27  RGADPOCFQO-----AR-----AVGLSTFAPFLSSRLQDLYSIVRRADRSV--P 70
DB  159  RGHDFACYQRTQLARQGLKVCVSHLIVGLPEGOA---ECLQTLERVVETGVGDKLHP 215
QY  71  IVNLKDEVLPSPWDSLFSGSGQLOLPGARI--FSPGDRDVLRRPAPQKSVWHGSDPSGR 128
DB  216  LHVKGSIWAKAWEA-----GRLN-GIELEDYTLTAGEMIHT--PPEVIYHRISASAR 266
QY  129  R---LMESYCET-W 138
DB  267  RPTLLAPLWCENRW 280

RESULT 9
SYJ1_HUMAN
ID   SYJ1_HUMAN          STANDARD;          PRT;   1575 AA.
AC   O43426; O43425;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-
DE   phosphate 1).
GN   SYNJ1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   TISSUE=Cerebellum;
RX   MEDLINE=98088905; PubMed=9428629;
RA   Haifner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,
RA   Salcini A.E., Di Fiore P.P., De Camilli P.;
RT   "Synaptotagmin 1: localization on coated endocytic intermediates in
RT   nerve terminals and interaction of its 170 kDa isoform with Eps15."
RL   FEBS Lett. 419:175-180(1997).
CC   -1- FUNCTION: inositol 5-phosphatase which has a role in clathrin-
CC   mediated endocytosis.
CC   -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC   biphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate
CC   + phosphate.
CC   -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1; Synonyms=Synaptotagmin-170;
CC   IsoId=O43426-1; Sequence=Displayed;
CC   Name=2; Synonyms=Synaptotagmin-145;
CC   IsoId=O43426-2; Sequence=VSP_002682; VSP_002683;
CC   -1- TISSUE SPECIFICITY: Concentrated at clathrin-coated endocytic
CC   intermediates in nerve terminals. Isoform 1 is more enriched than
CC   isoform 2 in developing brain as well as non-neuronal cells.
CC   -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
CC   C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.
CC   -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
CC   VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,
CC   SH3P4 AND GRB2.
CC   -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-
CC   TRISPHOSPHATE 5-PROSPHATASE FAMILY.
CC   -1- SIMILARITY: Contains 1 SAC domain.
CC   -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.

```

```

Dr  765 PPQG--PRTRMFSAGPTGSASS 784
DR  InterPro: IPR005911; Cons_hypoth1212.
DR  InterPro: IPR006638; ELP3.
DR  Pfam: PF04055; Radical SAM; 1.
DR  SMART: SM00729; ELP3; 1.
DR  TIGRFAMs: TIGR01212; TIGR01212; 1.
KW  Hypothetical protein; Complete proteome.
FT  CONFLICT 24 25   KL -> NV (IN REF. 4).
SQ  SEQUENCE 309 AA; 34606 MW; 61B3187BB77CA19 CRC64;

Query Match      8.2%; Score 76; DB 1; Length 309;
Best Local Similarity 26.1%; Pred. No. 4.3;
Matches 35; Conservative 23; Mismatches 42; Indels 34; Gaps 10;

QY  27  RGADPOCFQO-----AR-----AVGLSTFAPFLSSRLQDLYSIVRRADRSV--P 70
DB  159  RGHDFACYQRTQLARQGLKVCVSHLIVGLPEGOA---ECLQTLERVVETGVGDKLHP 215
QY  71  IVNLKDEVLPSPWDSLFSGSGQLOLPGARI--FSPGDRDVLRRPAPQKSVWHGSDPSGR 128
DB  216  LHVKGSIWAKAWEA-----GRLN-GIELEDYTLTAGEMIHT--PPEVIYHRISASAR 266
QY  129  R---LMESYCET-W 138
DB  267  RPTLLAPLWCENRW 280

RESULT 9
SYJ1_HUMAN
ID   SYJ1_HUMAN          STANDARD;          PRT;   1575 AA.
AC   O43426; O43425;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-
DE   phosphate 1).
GN   SYNJ1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX   TISSUE=Cerebellum;
RX   MEDLINE=98088905; PubMed=9428629;
RA   Haifner C., Takei K., Chen H., Ringstad N., Hudson A., Butler M.H.,
RA   Salcini A.E., Di Fiore P.P., De Camilli P.;
RT   "Synaptotagmin 1: localization on coated endocytic intermediates in
RT   nerve terminals and interaction of its 170 kDa isoform with Eps15."
RL   FEBS Lett. 419:175-180(1997).
CC   -1- FUNCTION: inositol 5-phosphatase which has a role in clathrin-
CC   mediated endocytosis.
CC   -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC   biphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate
CC   + phosphate.
CC   -1- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1; Synonyms=Synaptotagmin-170;
CC   IsoId=O43426-1; Sequence=Displayed;
CC   Name=2; Synonyms=Synaptotagmin-145;
CC   IsoId=O43426-2; Sequence=VSP_002682; VSP_002683;
CC   -1- TISSUE SPECIFICITY: Concentrated at clathrin-coated endocytic
CC   intermediates in nerve terminals. Isoform 1 is more enriched than
CC   isoform 2 in developing brain as well as non-neuronal cells.
CC   -1- DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A
CC   C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.
CC   -1- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
CC   VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,
CC   SH3P4 AND GRB2.
CC   -1- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-
CC   TRISPHOSPHATE 5-PROSPHATASE FAMILY.
CC   -1- SIMILARITY: Contains 1 SAC domain.
CC   -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF009040; AAC51922.1; -;
DR EMBL; AF009039; AAC51921.1; -;
DR GENE; HGNC:11503; SYNJ1.
DR MIM; 604297; -;
DR GO; GO:0005803; C:secretory vesicle; TAS.
DR GO; GO:0004445; F:inositol-1,4,5-trisphosphate 5-phosphatase . . .; TAS.
DR GO; GO:0006899; P:non-selective vesicle transport; TAS.
DR GO; GO:0006796; P:phosphate metabolism; TAS.
DR GO; GO:0008039; P:synaptic vesicle endocytosis; TAS.
DR InterPro; IPR00135; Exo_endo_phos.
DR InterPro; IPR000300; IPPC.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR002013; Syja_N.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF02383; Syja_N; 1.
DR SMART; SM00128; IPPC; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS0275; SAC; 1.
KW Hydrolase; Alternative splicing; Repeat; Endocytosis; RNA-binding;
KW Multigene family.
FT DOMAIN 119 442
FT DOMAIN 500 899
FT DOMAIN 902 971
FT DOMAIN 900 1575
FT DOMAIN 1033 1036
FT DOMAIN 1108 1113
FT DOMAIN 1126 1129
FT DOMAIN 1487 1490
FT DOMAIN 1540 1546
FT DOMAIN 1396 1419
FT REPEAT 1396 1398
FT REPEAT 1406 1408
FT REPEAT 1417 1419
FT VARSPLIC 1306 1311
FT VARSPLIC 1312 1575
FT VARSPLIC 1575 AA; 173345 MW; 50846 F6CC043B9E7 CRC64;
FT SEQUENCE 1575 AA; 173345 MW; 50846 F6CC043B9E7 CRC64;
Query Match 8.2%; Score 76; DB 1; Length 1575;
Best Local Similarity 22.6%; Pred. No. 31;
Matches 38; Conservative 21; Mismatches 49; Indels 60; Gaps 8;
QY 8 PVLHVALNTPLS--GGMRGIRG-----DFQCFOQAR-AVGLSGTFRAPLSRLQ 55
Db 648 PFIRDVAVDTVKTGMGGATGNKGAIVRMLEPHITSLCLFVCSHFAAGSQ-----VKERNE 702
QY 56 DLYSIVRRAD-----RGSPVIVNLKDEVLSWDSLFSGSQ--- 91
Db 703 DFIEIARKLSFPMGMLFSDHYVFCWGFNFYRIDLPNEEVKELLRQQNWDLSIAGDQLIN 762
QY 92 ----GQLQPG-----ARIFS--FDGRDVLHPANPQSVW 120
Db 763 QKNAGQVFRGLEKGVTFAPTYKYDLSFDSDYDTSEKCRTPAWTDRLVW 810
RESULT 10
CATA_METBA STANDARD; PRT; 505 AA.
AC 093662;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Catalase (EC 1.11.1.6).

GN KAT.
OS Methanosarcina barkeri.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2208;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fusaro / DSM 804;
RX MEDLINE=99311171; PubMed=10382262;
RA Shima S., Netrusov A., Sordel M., Wicke M., Hartmann G.C.,
RA Thauer R.K.;
RT "purification, characterization, and primary structure of a
RT monofunctional catalase from Methanosarcina barkeri";
RL Arch. Microbiol. 171:317-323(1999).
CC -!- FUNCTION: Decomposes hydrogen peroxide into water and oxygen;
CC serves to protect cells from the toxic effects of hydrogen
CC peroxide.
CC -!- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.
CC -!- COFACTOR: HEME GROUP (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the catalase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005939; CAA06774.1; -;
DR HSSP; P04040; 1F4J.
DR InterPro; IPR002226; Catalase.
DR Pfam; PF00199; Catalase; 1.
DR PRINTS; PR00067; CATALASE.
DR ProDom; PD000510; Catalase; 1.
DR PROSITE; PS00437; CATALASE 1; 1.
DR PROSITE; PS00438; CATALASE 2; 1.
KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 58 58
FT ACT_SITE 131 131 BY SIMILARITY.
FT METAL 341 341 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT SEQUENCE 505 AA; 57065 MW; 2A27C4BEC47B854 CRC64;
Query Match 8.1%; Score 75; DB 1; Length 505;
Best Local Similarity 20.7%; Pred. No. 9.8;
Matches 30; Conservative 30; Mismatches 47; Indels 38; Gaps 7;
QY 35 QQARAVGLSGTFRAPLSRLQDLYSIVRRADRGSPVIVNLKDEVLSWDSLFSGSQGL 94
Db 231 EEAEKIGGSDPDHA-----TRDLYEAIK---KGDYPSWTLEMQIWTPE----- 270
QY 95 QFCARIFSPDGRDVLHPANPQSVVHGSDDPS---GRRIMESYCYETWRTTGTGATQASS 151
Db 271 --QAEYRDIRDINK--VWP-----HGDFPTMKIGKLVLRNPTNYFAVEQAAAFSPAN 321
QY 152 LLSG-----RLLEKKAASCHNSYI 170
Db 322 LVFGIGISDPKMLQGRVFSYHDTHI 346
RESULT 11
SYNJ1_BOVIN STANDARD; PRT; 1324 AA.
ID SYNJ1_BOVIN
AC 018964;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Synaptotagmin 1 (EC 3.1.3.36) (Synaptic inositol-1,4,5-trisphosphate 5-
DE phosphatase 1) (p150) (Fragment).
GN SYNJ1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 321-339 AND 454-469.
 RC TISSUE=Brain;
 RX MEDLINE=97342621; PubMed=9199318;
 RA Sakisaka T., Itoh T., Miura K., Takenawa T.;
 RT "Phosphatidylinositol 4,5-bisphosphate phosphatase regulates the
 RT rearrangement of actin filaments";
 RL Mol. Cell. Biol. 17:3841-3849(1997).
 CC -!- FUNCTION: Hydrolyzes PIP2 bound to actin regulatory proteins
 CC resulting in the rearrangement of actin filaments downstream of
 CC tyrosine kinase and ASH/GRB2.
 CC -!- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
 CC bisphosphate + H(2)O = 1-phosphatidyl-1D-myo-inositol 4-phosphate
 CC + phosphate.
 CC -!- SUBUNIT: BINDS TO AMPHIPHYLIN AND ASH/GRB2.
 CC -!- SUBCELLULAR LOCATION: PREDOMINANTLY CONCENTRATED IN THE
 CC PERINUCLEAR AREAS.
 CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED WITH HIGHEST LEVELS IN
 CC BRAIN.
 CC -!- DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYLIN,
 CC AND ASH/GRB2.
 CC -!- SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-
 CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.
 CC -!- SIMILARITY: Contains 1 SAC domain.
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATIONS; A NUMBER OF POTENTIAL
 CC FRAMESHIFTS WERE CORRECTED STARTING IN POSITION 1213 SO AS TO
 CC EXTEND THE SIMILARITY WITH THE ORTHOLOGS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D85682; BAA21652.1; ALT_FRAME.
 DR InterPro; IPR005135; Exo_endo_phos.
 DR InterPro; IPR003000; IPFC.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR002013; SVIA_N.
 DR Pfam; PF03372; Exo_endo_phos; 1.
 DR Pfam; PF02383; Syjia_N; 1.
 DR SMART; SM00128; IPFC; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS50275; SAC; 1.
 DR Hydrolase; EC3.1.3.1; RNA-binding; Multigene family.
 KW Hydrolase; Endocytosis; RNA-binding; Multigene family.
 FT DOMAIN 119 442
 FT DOMAIN 475 859
 FT DOMAIN 902 971
 FT DOMAIN 860 1212
 FT DOMAIN 1033 1036
 FT DOMAIN 1108 1113
 FT DOMAIN 1126 1129
 FT DOMAIN 335 335
 FT CONFLICT 1324 1324
 FT NON_TER 1324 1324
 SQ SEQUENCE 1324 AA; 145582 MW; EDDC2DD9D6E3191C CRC64;
 Query Match 8.1%; Score 75; DB 1; Length 1324;
 Best Local Similarity 22.6%; Pred. No. 31;
 Matches 38; Conservative 20; Mismatches 50; Indels 60; Gaps 8;
 QY 8 PVLHLVALNTPLS--GGMEGIRGA-----DFQCFCQAR-AVGLSGTFRFLSSRLQ 55
 DB 648 PFIRDAVDVTVMKMGATGNKGVAIRMLFHTTSLCFTVCSEFAAGQSQ-----VKEND 702
 QY 56 DLYSIVRRAD-----RGSPVIVNLKDEVLSPSWDSLFSGSQ--- 91
 DB 703 DFLFIATKLSPFMGRLLFSDHVFVWCGDNFYIDLPNEEVKELIRQONWDSLIAGDQLIN 762

QY 92 ----GOLQPG-----ARIFS--FDGRDVLRHPAWPKSVW 120
 DB 763 QKNAGQIFRGFLGKVTFTPTKYDLSDDYDTSEKRTPAWTDRLVM 810
 RESULT 12
 RIAB CVHSA
 ID RIAB CVHSA STANDARD; PRT; 7073 AA.
 AC P59641;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Replicase polyprotein lab (pp1ab) [ORF1ab] [includes: Replicase
 DE polyprotein 1a (pp1a) [ORF1a]] [Contains: Leader protein; p65 homolog;
 DE Papain-like proteinase (EC 3.4.24.-) (NSP1); 3C-like proteinase
 DE (EC 3.4.24.-) (3CL-PRO) (NSP2); HD2 (NSP3); NSP4; NSP5; NSP6; Growth
 DE factor-like (NSP7); RNA-directed RNA polymerase (EC 2.7.7.48) (RDRP)
 DE (NSP9); Helicase (Hel) (NSP10); NSP11; NSP12; NSP13].
 OS Human coronavirus (strain SARS) (HCoV-SARS)
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Coronaviridae; Coronavirus.
 OX NCBI_TaxID=227859;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate Urbani;
 RA Bellini W.J., Campagnoli R.P., Icenogle J.P., Monroe S.S., Nix W.A.,
 RA Oberste M.S., Fallanisch M.A., Rota P.A.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate Tor2;
 RA Marra M., Jones S.J.M., Holt R.;
 RT "The complete genome of the SARS associated coronavirus";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate CUHK-WI;
 RA Tsui S.K.W., Lo D.Y.M., Tam J.S., Fung K.P., Chim S.S.C., Au C.C.,
 RA Chan A.H., Wan A.W.K., Au K.W., Chan C.W., Kou C.Y.C., Lam H.M.,
 RA Lam W.Y., Lau S.K., Lau Y.L., Lau Y.M., Law S.L., Law T.W., Li M.L.Y.,
 RA Tee C.H., Wong C.H., Yiu W.H., Lee C.Y., Chan A.K.C., Chiu R.W.K.,
 RA Ng E.K.O., Tong Y.K., Chan P.K.S., Au-Yeung C., Cheung J.K.L., Chu I.,
 RA Hung E.C.W., Waye M.M.Y.;
 RT "DNA sequence of a human coronavirus (CUHK-WI) from a patient with
 RT severe acute respiratory syndrome (SARS) in Hong Kong";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate HKU-39849;
 RA Leung F.C., Zeng F., Chan C.W.M., Chan C.M.Y., Chen J., Chow K.Y.C.,
 RA Hon C.C.C., Hui R.K.H., Li J., Li V.Y.Y., Wang Y.Y., Peiris J.S.M.,
 RA Poon L.L.M.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 4993-5127 FROM N.A.
 RC STRAIN=Isolate Vietnam;
 RA Emery S., Erdman D., Peret T., Ksiazek T.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 4993-5136 FROM N.A.
 RC STRAIN=Isolate Taiwan;
 RA Lin J.-H., Chiu S.-C., Yang J.-Y., Wang S.-F., Chen H.-Y.;
 RT "Detection of a novel human coronavirus in a severe acute respiratory
 RT syndrome patient in Taiwan";
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: The replicase polyprotein of coronaviruses is a
 CC multifunctional protein: it contains the activities necessary for
 CC the transcription of negative stranded RNA, leader RNA, subgenomic
 CC mRNAs and progeny virion RNA as well as proteinases responsible
 CC for the cleavage of the polyprotein into functional products (By
 CC similarity).
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

[RNA] (N).
-!- PNM: Specific enzymatic cleavages in vivo yield mature proteins
(By similarity).
-!- MISCELLANEOUS: This protein is translated as a 1A-1B polypeptide
by a ribosomal frameshifting mechanism (By similarity).
-!- SIMILARITY: Contains 1 peptidase family C16 domain.
-!- SIMILARITY: Contains 1 peptidase family C30 domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AY278741; AAP13442.1; -
EMBL; AY278741; AAP13439.1; -
EMBL; AY278741; AAP13440.1; ALT SEQ.
EMBL; AY274119; -; NOT_ANNOTATED_CDS.
EMBL; AY278554; AAP13566.1; -
EMBL; AY278554; AAP13575.1; -
EMBL; AY278491; -; NOT_ANNOTATED_CDS.
EMBL; AY269391; AAP04003.1; -
EMBL; AY268049; AAP04587.1; -
InterPro; IPR002589; ALPP.
InterPro; IPR007095; RNA_pol_DS_PS.
InterPro; IPR007094; RNA_pol_PSVir.
InterPro; IPR002877; FtsJ.
Pfam; PF01661; ALPP; 1.
Pfam; PF01728; FtsJ; 1.
SMART; SM00506; ALPP; 1.
PolyProtein; Transferrase; RNA-directed RNA polymerase; Thiol protease;
Hydrolase; Helicase; ATP-binding.
DOMAIN 1 179 LEADER PROTEIN (POTENTIAL).
DOMAIN 180 818 P65 HOMOLOG (POTENTIAL).
DOMAIN ? ? PAPAIN-LIKE PROTEINASE (POTENTIAL).
DOMAIN 3240 3547 3C-LIKE PROTEINASE (POTENTIAL).
DOMAIN 3548 3836 HD2/NSP3 (POTENTIAL).
DOMAIN 3837 3919 NSP4 (POTENTIAL).
DOMAIN 3920 4117 NSP5 (POTENTIAL).
DOMAIN 4118 4229 NSP6 (POTENTIAL).
DOMAIN 4230 4369 GROWTH FACTOR-LIKE (POTENTIAL).
DOMAIN 4370 5301 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
DOMAIN 5302 5902 HELICASE (POTENTIAL).
DOMAIN 5903 6429 NSP11 (POTENTIAL).
DOMAIN 6430 6775 NSP12 (POTENTIAL).
DOMAIN 6776 7073 NSP13 (POTENTIAL).
ACT SITE 1909 1909 ATP (POTENTIAL).
NP_BIND 5583 5590 ATP (POTENTIAL).
DOMAIN 930 933 POLY-GLU.
DOMAIN 937 942 POLY-GLU.
DOMAIN 974 979 POLY-GLU.
DOMAIN 2210 2213 POLY-LEU.
DOMAIN 3766 3769 POLY-CYS.
DOMAIN 2552 V -> A (in isolates Tor2, CUHK-W1 and HKU-39849).
DOMAIN 2556 D -> N (in isolate HKU-39849).
DOMAIN 2708 S -> T (in isolate HKU-39849).
DOMAIN 2718 R -> T (in isolate HKU-39849).
DOMAIN 3047 V -> A (in isolate CUHK-W1).
DOMAIN 3072 V -> A (in isolate CUHK-W1).
DOMAIN 3072 RVCG -> GFAY (in ORF1A).
DOMAIN 4382 CY -> VL (in isolate Taiwan).
DOMAIN 5131 A -> G (in isolate Taiwan).
DOMAIN 5135 D -> E (in isolate CUHK-W1).
DOMAIN 5767 Q -> R (in isolate Tor2).
DOMAIN 6778 D -> Y (in isolate Tor2).
DOMAIN 6883 D -> Y (in isolate Tor2).
DOMAIN 7073 AA; 790270 MW; A91B3CE920E69D4C CRC64;
SEQUENCE 7073 AA; 790270 MW; A91B3CE920E69D4C CRC64;
Query Match 8.1%; Score 75; DB 1; Length 7073;
Best Local Similarity 21.6%; Pred. No. 2.3e+02;
Matches 33; Conservative 25; Mismatches 65; Indels 30; Gaps 4;

QY 11 HLVALN-----TPLSGMGRGADFOCFQOARAVGLSGTFRALSLRLQDLYSIVRRA 64
DB 1623 YMSALNHTTKWKFPQVGLTSTIKWADNNCYLSSVLLAQLEVKFNAPALCEAYYRAG 1682
QY 65 DRGSVPINVLKDEVLSPSWDSLPSGSOQLOPGARIISFDFGRDLVRHPANPKSWHGSD 124
DB 1683 DAANFCALIL-----AYSNTKTVGELG-----DVRETMTW-----LLQHANL 1718
QY 125 PSGRRLMESYCETWRTTETTGATQASSLLSGL 157
DB 1719 ESAKRVNLVCKGCKQTITLTGVEAVMYMGT 1751
RESULT 13
PHBC RHIME
ID PHBC RHIME STANDARD; PRT; 611 AA.
AC P50176; O87321;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Poly-beta-hydroxybutyrate polymerase (EC 2.3.1.-) (Poly(3-
DE hydroxybutyrate) polymerase) (PHB polymerase) (PHB synthase) (Poly(3-
DE hydroxyalkanoate) polymerase) (PHA polymerase) (PHA synthase)
DE (Polyhydroxyalkanoic acid synthase).
GN PHBC OR R01725 OR SMC00296.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=41;
RX MEDLINE=96036213; PubMed=7582015;
RA Tomcolini R., Povolio S., Buson A., Squartini A., Nuti M.P.;
RT "Poly-beta-hydroxybutyrate (PHB) biosynthetic genes in Rhizobium
meliloti 41.";
RL Microbiology 141:2553-2559 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=98405222; PubMed=9734305;
RA Willis L.B., Walker G.C.;
RT "The phbC (poly-beta-hydroxybutyrate synthase) gene of Rhizobium
(Sinorhizobium) meliloti and characterization of phbC mutants.";
RL Can. J. Microbiol. 44:554-564 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramperger U.,
RA Renard C., Thebault P., Vandebol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
CC -!- FUNCTION: POLYMERIZES D(-)-3-HYDROXYBUTYRYL-COA TO CREATE PHB
WHICH CONSISTS OF THOUSANDS OF HYDROXYBUTYRATE MOLECULES LINKED
END TO END. PHB SERVES AS AN INTRACELLULAR ENERGY RESERVE MATERIAL
WHEN CELLS GROW UNDER CONDITIONS OF NUTRIENT LIMITATION.
CC -!- PATHWAY: Poly-beta-hydroxybutyrate biosynthesis; third step.
CC -!- SUBUNIT: Monomer (Probable).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; UI7227; AAA90984.1; ALT INIT.
DR EMBL; AF031938; AAC61899.1; -.
DR EMBL; AL591788; CAC46304.1; -.
KW PHB biosynthesis; Transferase; Acyltransferase; Complete proteome.
FT ACT_SITE 349 349 POTENTIAL.
FT CONFLICT 111 111 G -> A (IN REF. 1).
FT CONFLICT 122 122 R -> H (IN REF. 1).
FT CONFLICT 157 157 R -> K (IN REF. 1).
FT CONFLICT 477 477 K -> R (IN REF. 1).
FT CONFLICT 481 481 V -> M (IN REF. 1).
SQ SEQUENCE 611 AA; 68039 MW; ABE8A9B17F87D7A9 CRC64;

Query Match 8.1%; Score 74.5; DB 1; Length 611;
Best Local Similarity 27.4%; Pred. No. 14;
Matches 40; Conservative 16; Mismatches 47; Indels 43; Gaps 9;

QY 49 FLSSRLQ--DLYSIVRRDRGSV--PIVNL--KDEVLSPSWDSLFSGSQQLQPGARIFS 102
DB 470 YLENRLSKGEMVLAGRVSIGDVKIPIYNLATKEDHIAPA-KSVFLGSS-----S 518

QY 103 PDGR-----DVLRRPAPQKSVWHSPPSGRLMESYCETW---RRETTGA-- 145
DB 519 FGGKVTFTVLSGSHIAGVNPAPRSKYQYWTGGAPKGD-----IETWNGKAKETAGSWW 572

QY 146 ---TQASLLSGRLLEOKAASCHNS 168
DB 573 PHMQGWZELDKRRVPARKAGGPLNS 598

RESULT 14
P47K_PSECL STANDARD; PRT; 419 AA.
AC P31521;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 47 kDa protein (P47K).
OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=333;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=91193202; PubMed=2013568;
RA Nishiyama M., Hozinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA Beppu T.;
RT "Cloning and characterization of genes responsible for metabolism of
RT nitrile compounds from Pseudomonas chlororaphis B23.";
RL J. Bacteriol. 173:2465-2472(1991).
CC -!- FUNCTION: THE PRESENCE OF P47K IS CRITICAL FOR THE EXPRESSION
CC OF THE NITRILE HYDRATASE GENES. MAY STABILIZE OR ACTIVATE THE
CC NITRILE HYDRATASE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90216; BAA14247.1; -.
DR FIR; D42725; D42725.
DR InterPro; IPR003495; CobW.
DR Pfam; PF02492; CobW; 1.
SQ SEQUENCE 419 AA; 46666 MW; FF5113800E27FF0C CRC64;

Query Match 7.9%; Score 72.5; DB 1; Length 419;
Best Local Similarity 22.1%; Pred. No. 14;
Matches 40; Conservative 24; Mismatches 60; Indels 57; Gaps 8;

```

```

QY 26 IRGADPQEQQARAVLGSCTFFAFSSR-LQDLSIVRADRGSVPIVNLKDEVLSPSWD 84
DB 141 VDGSOQFALLESTDTVARADTEAHTSTRHLADL--LIQVEYANVILVKNRDLIDEPGQ 198
QY 85 SLFSGSQQLQPGARIFS-----FDGRDVLRRHPAPQK-----SVWHGS 123
DB 199 AVHAILAG-LNPSARIMPVAGHVALSSLLDTHLFDLPSLAASPGWMRKMEATDTPASES 257
QY 124 DPSG-----RRLMESYCETWRTTGTATGQASSLLSGRLLEOKA-----ASC 165
DB 258 DTGVTSTWYRERAPPHPQRLLEFLQKPMH-----NGLRLRSKGYFWLASR 303
QY 166 H 166
DB 304 H 304

RESULT 15
VNUC_EBOGA STANDARD; PRT; 739 AA.
ID VNUC_EBOGA
AC O9OC69;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Nucleoprotein (Nucleocapsid protein).
GN NP.
OS Ebola virus (strain Gabon-94) (Ebol).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC Ebola-like viruses.
OX NCBI_TaxID=128947;
RN [1];
RP SEQUENCE FROM N.A.
RX MEDLINE=99036017; PubMed=9820131;
RA Prehaud C.J.C., Hellebrand E., Coudrier D., Volchkov V.E.,
RA Volchkova V.A., Feldmann B., Le Guenno B., Bouloy M.;
RT "Recombinant Ebola virus nucleoprotein and glycoprotein (Gabon 94
RT strain) provide new tools for the detection of human infections.";
RL J. Gen. Virol. 79:2585-2592(1998).
CC -!- FUNCTION: RESPONSIBLE FOR ENCAPSIDATION OF GENOMIC RNA.
CC -!- DOMAIN: THIS PROTEIN CAN BE DIVIDED INTO A HYDROPHOBIC N-TERMINAL
CC HALF, AND A HYDROPHILIC AND HIGHLY ACIDIC C-TERMINAL HALF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09358; CAA70541.1; -.
KW Nucleoprotein.
SQ SEQUENCE 739 AA; 83240 MW; 33A96720FFBC638E CRC64;

Query Match 7.9%; Score 72.5; DB 1; Length 739;
Best Local Similarity 27.0%; Pred. No. 27;
Matches 43; Conservative 16; Mismatches 67; Indels 33; Gaps 8;

QY 22 GMRGIGAD-----PQCQQRARVGLSGTFFAFSSRLQD-----LYSIVRRDRGSV 69
DB 218 GMMVAGHDANDAVINSVAQARFSGLL-IVKTVLDHILQKTQGVRLHPLARTA----- 271
QY 70 PIVNLKDEV--LSPSWDSLFSGSQQLQPGARIFSDGRDVLRRHPAPQK-----VW 120
DB 272 ---KVNENVSLKAALSSL--AKHGVEAPFARLLNLGYNLLEHGLFQPSALGAVATA 326
QY 121 HGSDDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLE 159
DB 327 HGSTLAGVNVGEQY-QQLFEAAATEAKQLQQAARESLD 364

Search completed: February 17, 2004, 09:57:04

```


Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 17, 2004, 09:48:08 ; Search time 35 Seconds
(without alignments)
1290.263 Million cell updates/sec

Title: US-09-589-777C-2_COPY_1_175

Perfect score: 923

Sequence: 1 HTHQDFQPVHLVALNTPLS.....RLLEQKAASCHNSYIVLCIE 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923	100.0	184	11 Q9JK63	Q9JK63 mus musculus
2	923	100.0	1140	11 Q61434	Q61434 mus musculus
3	923	100.0	1774	11 Q62001	Q62001 mus musculus
4	882	95.6	226	11 Q9Q2D2	Q9Q2D2 rattus norv
5	837	90.7	171	11 Q9WUW5	Q9WUW5 rattus norv
6	813	88.1	187	4 Q8WX15	Q8WX15 homo sapien
7	813	88.1	261	4 Q8NG19	Q8NG19 homo sapien
8	813	88.1	816	4 Q8N4S4	Q8N4S4 homo sapien
9	793	85.9	160	11 Q9CRT2	Q9CRT2 mus musculus
10	724	78.4	1344	13 Q93419	Q93419 gallus gall
11	692	75.0	1307	13 Q8JFF7	Q8JFF7 xenopus lae
12	690	74.8	1315	13 Q8QHL9	Q8QHL9 xenopus lae
13	605.5	65.6	361	13 Q8AWC6	Q8AWC6 brachydanio
14	546	59.2	1367	11 Q35206	Q35206 mus musculus
15	545	59.0	1367	11 Q9EQD9	Q9EQD9 mus musculus
16	536	58.1	1388	4 Q9Y4W4	Q9Y4W4 homo sapien

ALIGNMENTS

RESULT 1

Q9JK63 ID Q9JK63 PRELIMINARY; PRT; 184 AA.

AC Q9JK63;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Endostatin (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Chinese Kunming;
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;
RT "Anticancer treatment of targeted fusion protein delivery to tumor neovasculture."
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257775; AAF69009.1; -
DR HSSP; P39061; IKOE.
FT NON_TER 1
FT NON_TER 184 184
SQ SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;

Query Match 100.0%; Score 923; DB 11; Length 184;
Best Local Similarity 100.0%; Pred. No. 2.7e-84;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGGMRGTRGADFQCFQARAVGLSGTFRAFLSRLQDLYSI 60
DB 1 HTHQDFQPVHLVALNTPLSGGMRGTRGADFQCFQARAVGLSGTFRAFLSRLQDLYSI 60
QY 61 VRRADRGSPVPIVNLKDEVLSPSWDSILFSSGQQLQFGARIFSDGRDVLHPAPQKSVW 120
DB 61 VRRADRGSPVPIVNLKDEVLSPSWDSILFSSGQQLQFGARIFSDGRDVLHPAPQKSVW 120
QY 121 HSDPSGRRLMSYCYETWTETTGATGQASSLLSGLLEQKAAASCHNSYIVLCIE 175

Q96T70 homo sapien
Q8mse3 drosophila
Q8mt89 drosophila
O17866 caenorhabdi
Q9u9k6 caenorhabdi
Q9u9k7 caenorhabdi
Q810g6 caenorhabdi
Q92k28 rhizobium m
Q9snx8 streptomyce
Q8r863 leptospira
Q9pjg2 xanthomona
Q9vfa9 drosophila
Q914y1 streptomyce
Q8v0c4 ralstonia s
Q8fde5 escherichia
Q97f93 clostridium
Q82700 rattus norv
Q9yur6 turkey aden
Q93nrx9 streptomyce
Q82394 salmonella
Q816g9 salmonella
Q8zv30 pyrobaculum
Q914x3 streptomyce
Q9vsn3 drosophila
Q8n159 homo sapien
Q39725 euglena gra
Q8t0w6 echinococcu
Q8ghl6 pseudomonas
Q50185 mycobacteri

```

Db 121 HGSDFSGRRLMESYCEWTWTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
RESULT 2
Q61434
ID Q61434 PRELIMINARY; PRT; 1140 AA.
AC Q61434
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region.";
RL Cell. Mol. Biol. Res. 196;576-582(1993).
DR EMBL; D17546; BAA04483.1; -.
DR HSSP; P39061; IKOE.
DR MGD; MGI:88449; Coll15a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 1140 AA; 1 115156 MW; 8B0C7E6862B3BDFE CRC64;

Query Match 100.0%; Score 923; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 2.8e-83;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 60
DB 957 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 1016
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQGRDVLHRHPAPQKSVW 120
DB 1017 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQGRDVLHRHPAPQKSVW 1076
QY 121 HGSDFSGRRLMESYCEWTWTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
DB 1077 HGSDFSGRRLMESYCEWTWTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1131

RESULT 3
Q62001
ID Q62001 PRELIMINARY; PRT; 1774 AA.
AC Q62001; Q60672;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen
DE (Procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen

```

```

chain.";
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M., Pihlajaniemi T.;
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RX MEDLINE=95181468; PubMed=7876242;
RA Rehn M., Pihlajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding
RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins.";
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL; U03715; AAC52903.1; JOINED.
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.
DR EMBL; U34609; AAC52903.1; JOINED.
DR EMBL; U34610; AAC52903.1; JOINED.
DR EMBL; U34611; AAC52903.1; JOINED.
DR EMBL; U34612; AAC52903.1; JOINED.
DR EMBL; U34613; AAC52903.1; JOINED.
DR EMBL; U11637; AAC52179.1; -.
DR HSSP; P39061; IKOE.
DR MGD; MGI:88451; Coll18a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000024; Fz domain.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00063; FRI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS50038; FZ; 1.
KW Collagen; Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 923; DB 11; Length 1774;
Best Local Similarity 100.0%; Pred. No. 5e-83;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 60
DB 1591 HTHQDFQPVHLVALNTPLSGMRGIRGADFQCQQAQAVGLSGTFFRAFLSSRLQDLYSI 1650
QY 61 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQGRDVLHRHPAPQKSVW 120
DB 1651 VRRADRGSPVIVNLKDEVLSFSGSQQLQPGARIFSFQGRDVLHRHPAPQKSVW 1710
QY 121 HGSDFSGRRLMESYCEWTWTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
DB 1711 HGSDFSGRRLMESYCEWTWTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1765

RESULT 4
Q9QZD2
ID Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Collagen XVIII (Fragment).
OS Rattus norvegicus (Rat).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RX MEDLINE=2027226; PubMed=10766159;
 RA Parletti G., Concari P., Giardini R., Marras E., Piccinini F.,
 RA Folkman J., Chen L.;
 RT "Antitumor activity of endostatin against carcinogen-induced rat
 RT primary mammary tumors.";
 RL Cancer Res. 60:1793-1796(2000).
 DR EMBL; AF189709; AA00975.1; -.
 DR HSP; P39061; IKOE.
 FT NON_TER 1
 SQ SEQUENCE 226 AA; 23350 MW; 38B93C0486C0E949 CRC64;
 Query Match 95.6%; Score 882; DB 11; Length 226;
 Best Local Similarity 95.4%; Pred. No. 4.5e-80;
 Matches 167; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 1 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
 DB 43 HTHQDFHVLHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 102
 QY 61 VRRADRGSPVIVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVW 120
 DB 103 VRRADRGSPVIVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVW 162
 QY 121 HGSPPSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 175
 DB 163 HGSPPSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 217
 RESULT 5
 Q9WUW5
 ID Q9WUW5 PRELIMINARY; PRT; 171 AA.
 AC Q9WUW5;
 DT 01-NOV-1999 (TremBLrel. 12, Created)
 DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
 DE Collagen type XVIII, alpha (I) chain (Fragment).
 GN COL18A1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Jia J.D., Bauer M., Eberpaecher U., Donner P., Schuppan D.;
 RA "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;
 RA "temporospatial expression of collagen XVIII/endostatin in acute and
 RT chronic liver injuries.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ236873; CAB44263.1; -.
 DR HSP; P39061; IKOE.
 FT NON_TER 1
 SQ SEQUENCE 171 AA; 19933 MW; 81B2E3FC2C8E72 CRC64;
 Query Match 90.7%; Score 837; DB 11; Length 171;
 Best Local Similarity 95.8%; Pred. No. 9.8e-76;
 Matches 160; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 9 VHLHVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSIVRRDRGS 68
 DB 1 VHLHVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSIVRRDRGS 60

QY 69 VPVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVWHGSDPSGR 128
 DB 61 VPVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVWHGSDPSGR 120
 QY 129 RLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 175
 DB 131 RLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 167

RESULT 6

Q8WXI5

ID Q8WXI5 PRELIMINARY; PRT; 187 AA.
 AC Q8WXI5;
 DT 01-WAR-2002 (TremBLrel. 20, Created)
 DT 01-WAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-WAR-2002 (TremBLrel. 20, Last annotation update)
 DE Collagen XVIII (Fragment).
 OS Homo sapiens (Human).
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21409408; PubMed=11517600;
 RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
 RT "Inhibition effect in vitro of purified endostatin expressed in Pichia
 RT pastoris.";
 RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
 DR EMBL; AF416592; AAL37720.1; -.
 FT NON_TER 1
 SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85838CD3 CRC64;

Query Match

Best Local Similarity 88.1%; Score 813; DB 4; Length 187;
 Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHQDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 60
 DB 5 HSHRDFQVHLVALNTPLSGMRGIRGADFCQFQARAVGLSGTFRFLSSRLQDLYSI 64
 QY 61 VRRADRGSPVIVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVW 120
 DB 65 VRRADRGSPVIVNLKDEVLSFSGSQGLHSGARIFSPDGRDVLHRHPAWPKQSVW 124
 QY 121 HGSPPSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 175
 DB 125 HGSPPSGRRRLMESYCETWRTTGTATGQASSLLSGRLLEKQAACHNSYIVLCIE 179

RESULT 7

Q8NGI9

ID Q8NGI9 PRELIMINARY; PRT; 261 AA.
 AC Q8NGI9;
 DT 01-OCT-2002 (TremBLrel. 22, Created)
 DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
 DE Multi-functional protein MFP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Dou D.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF282883; AAMS2249.1; -.
 SQ SEQUENCE 261 AA; 26745 MW; CA60C920AF3B90E5 CRC64;

Query Match

Best Local Similarity 88.1%; Score 813; DB 4; Length 261;
 Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

```
QY 1 HTHODFQVHLVALNTPISGMRGIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSI 60
Db 79 HSHRDFQVHLVALNTPISGMRGIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSI 138
QY 61 VRRADRGVPIVNLKDEVLSPSWDSLFSGQQLQPCGARIFFSGDRDVLHRHPAWPKSVW 120
Db 139 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSGDKVLRHPTWPKSVW 198
QY 121 HGSDDSGRLMESYCETWRTTGTATGQASSLLSGLLEKQAACHNSYIVLCIE 175
Db 199 HGSDDNGRLTESYCETWRTTGTATGQASSLLSGLLEKQAACHNSYIVLCIE 253

RESULT 8
Q8N4S4 PRELIMINARY; PRT; 816 AA.
AC Q8N4S4;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Similar to collagen, type XVIII, alpha 1 (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Renal adenocarcinoma;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033715; AAH33715.1; -.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Collagen; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;

Query Match 88.1%; Score 813; DB 4; Length 816;
Best Local Similarity 85.7%; Pred. No. 1.8e-72;
Matches 150; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 HTHODFQVHLVALNTPISGMRGIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSI 60
Db 634 HSHRDFQVHLVALNTPISGMRGIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSI 693
QY 61 VRRADRGVPIVNLKDEVLSPSWDSLFSGQQLQPCGARIFFSGDRDVLHRHPAWPKSVW 120
Db 694 VRRADRAAVPIVNLKDELLFSSWEALFSGSEGLPKPGARIFSGDKVLRHPTWPKSVW 753
QY 121 HGSDDSGRLMESYCETWRTTGTATGQASSLLSGLLEKQAACHNSYIVLCIE 175
Db 754 HGSDDNGRLTESYCETWRTTGTATGQASSLLSGLLEKQAACHNSYIVLCIE 808

RESULT 9
Q8C72 PRELIMINARY; PRT; 160 AA.
AC Q8C72;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Procollagen, type XVIII, alpha 1 (Fragment).
GN COL18A1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX STRAIN=C57BL/6J; PubMed=11217851;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
```

```
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014292; BAB29249.1; -.
DR HSP; P39061; IKOE.
DR MGD; MGI:89451; Coll18a1.
FT NON TER
SQ SEQUENCE 160 AA; 17725 MW; 60F853D777C375D2 CRC64;

Query Match 85.9%; Score 793; DB 11; Length 160;
Best Local Similarity 100.0%; Pred. No. 2.2e-71;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 GIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSIIVRRADRGVPIVNLKDEVLSPSWD 84
Db 1 GIRGADFCQFQARAVGLSCTFFAFSSRLQDLYSIIVRRADRGVPIVNLKDEVLSPSWD 60
QY 85 SLFSGSQQLQPCGARIFFSGDRDVLHRHPAWPKSVVHSGDSPGRRLMESYCETWRTTGT 144
Db 61 SLFSGSQQLQPCGARIFFSGDRDVLHRHPAWPKSVVHSGDSPGRRLMESYCETWRTTGT 120
QY 145 ATQASSLLSGLLEKQAACHNSYIVLCIE 175
Db 121 ATQASSLLSGLLEKQAACHNSYIVLCIE 151

RESULT 10
Q93419 PRELIMINARY; PRT; 1344 AA.
AC Q93419;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE Collagen XVIII precursor.
OS Gallus gallus (Chicken)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411346; PubMed=9738008;
RA Halfter W., Dong S., Schurer B., Cole G.J.;
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 273:25404-25412(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083440; AAC33294.2; -.
DR HSP; P39061; IKOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 2.
```

```

DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Signal. 26
FT SIGNAL 1
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;

Query Match 78.4%; Score 724; DB 13; Length 1344;
Best Local Similarity 76.0%; Pred. No. 2.7e-63;
Matches 133; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 60
DB 1161 HVHQDFQFALHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 1220
QY 61 VRRADRGSVPIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 120
DB 1221 VRRADRTAVPIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 1280
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
DB 1281 HGSDAKGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1335

RESULT 11
Q8JFF7 PRELIMINARY; PRT; 1307 AA.
AC Q8JFF7
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XVIII collagen short variant.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2216979; PubMed=12175494;
RA Elamra H., Peterson J., Pihlajaniemi T., Destree O.;
RT "Cloning of three variants of type XVIII collagen and their expression
patterns during Xenopus laevis development.";
RL Mech. Dev. 114:109-113(2002).
DR EMBL; AY052763; AAL14257.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1307 AA; 134153 MW; D53EDBPE3DE34976 CRC64;

Query Match 75.0%; Score 692; DB 13; Length 1307;
Best Local Similarity 73.1%; Pred. No. 4.1e-60;
Matches 128; Conservative 16; Mismatches 31; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 60
DB 1124 HTHQDFQFALHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 1183
QY 61 VRRADRGSVPIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 120
DB 1184 VRRADRGSVQIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 1243
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
DB 1244 HGSDAKGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1298

RESULT 12
Q8QHL9 PRELIMINARY; PRT; 1315 AA.
AC Q8QHL9
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XVIII collagen alpha1 chain.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2216979; PubMed=12175494;
RA Ishino T., Sekimizu K., Natori S., Kubo T.;
RT "Identification and characterization of genes expressed selectively in
the regenerating tail of Xenopus laevis tadpole.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047066; BAB84674.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 7.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1315 AA; 134946 MW; OC56C235DE058365 CRC64;

Query Match 74.8%; Score 690; DB 13; Length 1315;
Best Local Similarity 73.1%; Pred. No. 6.6e-60;
Matches 128; Conservative 16; Mismatches 32; Indels 0; Gaps 0;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 60
DB 1132 HTHQDFQFALHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 1191
QY 61 VRRADRGSVPIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 120
DB 1192 VRRADRGSVQIVNLKDEVLSFSGSQGLQPCGARIFFSDGRDVLHHPAPQKSVW 1251
QY 121 HGSDPSGRRLMESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 175
DB 1252 HGSDAKGRRLTESYCETWRTTGTATGQASSLLSGRLLEQKAASCHNSYIVLCIE 1306

RESULT 13
Q8AWC6 PRELIMINARY; PRT; 361 AA.
AC Q8AWC6
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Collagen XVIII (Fragment).
GN COL18A1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteiophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2216979; PubMed=12175494;
RA Haftek Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT "Sequence and embryonic expression of collagen XVIII NC11 domain
(endostatin) in the zebrafish.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ494837; CAD38825.1; -.
DR SMART; SM00210; TSPN; 1.
KW NON_TER
SQ SEQUENCE 361 AA; 40222 MW; 3CSA0F8479D26735 CRC64;

Query Match 65.6%; Score 605.5; DB 13; Length 361;
Best Local Similarity 65.1%; Pred. No. 3.5e-52;
Matches 114; Conservative 24; Mismatches 24; Indels 1; Gaps 1;

QY 1 HTHQDFQPVHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 60
DB 179 HTHQDFQFALHLVALNTPLSGMRGIRGADFCQFOQARAVLSGTFFAFSLSSRLQDLYSI 237

```

```

QY 61 VRRADRGVPIVNLKDEVLSPSWDSLPSGSGQLQPGARIISFGDRDVLHPAPQKSVW 120
DB 238 VRRSDRELPIVNLKQDLFWSWESLFSRSRKDNAPISFGDRDVLDSAPWPKVW 297
QY 121 HGSDFSGRRLMESYCEWTWTTGATGQASLLSGRLLEQKAAACHNSYIVLCIE 175
DB 298 HGSDFSGRRQTDNYCETWAGDRAVTVGLASSLQAGQLLQQTSSSCSSSYIALCIE 352

RESULT 14
O35206 PRELIMINARY; PRT; 1367 AA.
AC O35206;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 23, Last annotation update)
DE Type XV collagen.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains.";
RL Genomics 45:31-41(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=129/Sv;
RX MEDLINE=20522048; PubMed=11068203;
RA Eklund L., Wuona A., Lietaud J., Pihlajaniemi T.;
RT "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT with the human COL15A1 gene and functional analysis of the promoters
RT of both genes.";
RL Matrix Biol. 19:489-500(2000).
DR EMBL; AF261131; AAG27545.1; JOINED.
DR EMBL; AF261109; AAG27545.1; JOINED.
DR EMBL; AF261110; AAG27545.1; JOINED.
DR EMBL; AF261111; AAG27545.1; JOINED.
DR EMBL; AF261112; AAG27545.1; JOINED.
DR EMBL; AF261113; AAG27545.1; JOINED.
DR EMBL; AF261114; AAG27545.1; JOINED.
DR EMBL; AF261115; AAG27545.1; JOINED.
DR EMBL; AF261116; AAG27545.1; JOINED.
DR EMBL; AF261117; AAG27545.1; JOINED.
DR EMBL; AF261118; AAG27545.1; JOINED.
DR EMBL; AF261119; AAG27545.1; JOINED.
DR EMBL; AF261120; AAG27545.1; JOINED.
DR EMBL; AF261121; AAG27545.1; JOINED.
DR EMBL; AF261122; AAG27545.1; JOINED.
DR EMBL; AF261123; AAG27545.1; JOINED.
DR EMBL; AF261124; AAG27545.1; JOINED.
DR EMBL; AF261125; AAG27545.1; JOINED.
DR EMBL; AF261126; AAG27545.1; JOINED.
DR EMBL; AF261127; AAG27545.1; JOINED.
DR EMBL; AF261128; AAG27545.1; JOINED.
DR EMBL; AF261129; AAG27545.1; JOINED.
DR EMBL; AF261130; AAG27545.1; JOINED.
DR HSSP; P39061; 1KOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEEC CRC64;

Query Match 59.2%; Score 546; DB 11; Length 1367;
Best Local Similarity 62.1%; Pred. No. 1.7e-45;
Matches 105; Conservative 22; Mismatches 38; Indels 4; Gaps 1;

QY 7 QPVLHLVALNTPLSGMGRGADFCQQAARAVGLSGTFFAFLSSRLQDLYSIVRRADR 66
DB 1194 RPYLHLVALNTFVAGDIR----ADFCQQAARAGLLSTFFAFLSSHLQDLSTVVRKAER 1249
QY 67 GSVPIVNLKDEVLSPSWDSLPSGSGQLQPGARIISFGDRDVLHPAPQKSVWHSNPH 126
DB 1250 FGLPIVNLKGVLFNNWDSIFSGDGGQNTHTPIYSFGDRDVTDPWPKVWHSNPH 1309

RESULT 15
Q9EQD9 PRELIMINARY; PRT; 1367 AA.
AC Q9EQD9;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Type XV collagen.
GN COL15A1.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=129/Sv;
RX MEDLINE=97480713; PubMed=9339358;
RA Hagg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains.";
RL Genomics 45:31-41(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=129/Sv;
RX MEDLINE=20522048; PubMed=11068203;
RA Eklund L., Wuona A., Lietaud J., Pihlajaniemi T.;
RT "Structure of the mouse type XV collagen gene, Col15a1, comparison
RT with the human COL15A1 gene and functional analysis of the promoters
RT of both genes.";
RL Matrix Biol. 19:489-500(2000).
DR EMBL; AF261131; AAG27545.1; JOINED.
DR EMBL; AF261109; AAG27545.1; JOINED.
DR EMBL; AF261110; AAG27545.1; JOINED.
DR EMBL; AF261111; AAG27545.1; JOINED.
DR EMBL; AF261112; AAG27545.1; JOINED.
DR EMBL; AF261113; AAG27545.1; JOINED.
DR EMBL; AF261114; AAG27545.1; JOINED.
DR EMBL; AF261115; AAG27545.1; JOINED.
DR EMBL; AF261116; AAG27545.1; JOINED.
DR EMBL; AF261117; AAG27545.1; JOINED.
DR EMBL; AF261118; AAG27545.1; JOINED.
DR EMBL; AF261119; AAG27545.1; JOINED.
DR EMBL; AF261120; AAG27545.1; JOINED.
DR EMBL; AF261121; AAG27545.1; JOINED.
DR EMBL; AF261122; AAG27545.1; JOINED.
DR EMBL; AF261123; AAG27545.1; JOINED.
DR EMBL; AF261124; AAG27545.1; JOINED.
DR EMBL; AF261125; AAG27545.1; JOINED.
DR EMBL; AF261126; AAG27545.1; JOINED.
DR EMBL; AF261127; AAG27545.1; JOINED.
DR EMBL; AF261128; AAG27545.1; JOINED.
DR EMBL; AF261129; AAG27545.1; JOINED.
DR EMBL; AF261130; AAG27545.1; JOINED.
DR HSSP; P39061; 1KOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;

Query Match 59.0%; Score 545; DB 11; Length 1367;
Best Local Similarity 62.1%; Pred. No. 2.1e-45;
Matches 105; Conservative 22; Mismatches 38; Indels 4; Gaps 1;

QY 7 QPVLHLVALNTPLSGMGRGADFCQQAARAVGLSGTFFAFLSSRLQDLYSIVRRADR 66
DB 1194 RPYLHLVALNTFVAGDIR----ADFCQQAARAGLLSTFFAFLSSHLQDLSTVVRKAER 1249
QY 67 GSVPIVNLKDEVLSPSWDSLPSGSGQLQPGARIISFGDRDVLHPAPQKSVWHSNPH 126
DB 1250 FGLPIVNLKGVLFNNWDSIFSGDGGQNTHTPIYSFGDRDVTDPWPKVWHSNPH 1309
QY 127 GRRLMESYCEWTWTTGATGQASLLSGRLLEQKAAACHNSYIVLCIE 175

```

Db 1310 GVLVDKYCEAWRTTDMVAVTGFAPISTGKILDQKAYSCANRLIVLCIE 1358

Search completed: February 17, 2004, 09:57:53
Job time : 38 secs